African swine fever pathogenesis: Comparative analysis of immunoregulatory genes in domestic and wild pigs

Christopher J. Palgrave

University of Edinburgh

Veterinary Biomedical Sciences

College of Medicine & Veterinary Medicine

Doctor of Philosophy

2004

THE UNIVERSITY OF EDINBURGH

ABSTRACT OF THESIS

(Regulation 3.9.14)

Name of Candidate: Christopher James Palgrave

Address: Veterinary Biomedical Sciences, Royal (Dick) School of Veterinary Studies, College of Medicine & Veterinary Medicine, University of Edinburgh, Summerhall, Edinburgh

Postal Code: EH9 1QH

Degree: Doctor of Philosophy

Title of Thesis: African swine fever pathogenesis: Comparative analysis of immunoregulatory genes in domestic and wild pigs.

No. of words in the main text of Thesis: 45 000

Abstract:

African swine fever virus (ASFV) poses one of the greatest threats to pig farming worldwide. It is highly infectious and causes rapid haemorrhagic death of domestic pigs and Eurasian wild boar (*Sus scrofa*). In contrast, native African pig species (bushpigs and warthogs) suffer only a mild, subclinical form of the infection from which they rapidly recover. It is hypothesised that the striking difference in the pathophysiological consequences of ASFV infection may reflect a variable ability of the virus to modulate the host immune response in these species. Alternatively, it may indicate a fundamental evolutionary distinction between the immune responses of these animals.

In common with many other DNA viruses, ASFV has evolved a complex strategy for modulating the host-cell immune response. The ASFV-encoded protein, A238L, targets key sites within both the NFκB and NFAT immune-signalling pathways. Furthermore, the ASFV protein, p54, is involved in attachment of virus particles to the microtubule motor complex, cytoplasmic dynein. This may represent a key stage in the infection process. Six host proteins targeted or mimicked by A238L and p54 (light chain dynein, cyclophilin A, calcineurin A, NFAT, p65 (RelA) and IκBα) have been sequenced in the susceptible domestic pig, resistant warthog and phenotypically unknown babirusa. In addition, the ~1.6kbp promoter driving expression of the proinflammatory cytokine, tumour necrosis factor alpha (TNFα), has also been studied.

Despite identifying high levels of nucleotide sequence conservation in these genes, polymorphisms have been identified in the NFκB subunit p65 (RelA) and the TNFα promoter. These may be of functional significance in determining the immune response characteristic of the different pig species studied. These polymorphisms have been further explored using *in vitro* expression and luciferase-reporter analysis. Furthermore, the identification of these sites has enabled the commercial sponsor of this project, Sygen International, to screen their domestic pig lines for 'warthog-like' sequence, which may confer some degree of disease resistance.

These findings provide a valuable insight into potential mechanisms involved in altered host susceptibility to African swine fever. In addition, this study may have wider-reaching implications for understanding issues of both susceptibility and pathogenesis relating to other infectious diseases of both humans and animals.

Acknowledgements

I would like to thank my supervisors, Dr. Stewart Lowden, Dr. Susan Rhind and Dr. Paul Hopwood, and my external advisor, Dr. Bruce Whitelaw, for their invaluable support, advice and guidance over the last three years.

I am particularly grateful to Susan Rhind who kindly agreed to take on the role of primary supervisor at short notice and Bruce Whitelaw who welcomed me into his group at the Roslin Institute.

Many thanks also to Dr. Marnie Mellencamp and Sygen International plc for both their practical and financial support for this work. This project was also generously funded by a Royal (Dick) School of Veterinary Studies Postgraduate Centenary Fellowship.

Declaration

I declare that the composition of this thesis and the work presented herein are my own, except where specifically stated.

Christopher Palgrave

Table of Contents

List of Figures

List of Tables and Information Boxes

Abbreviations

ASF	African swine fever
ASFV	African swine fever virus
BB	*Babyrousa babyrussa* (Babirusa)
BCA	Bicinchoninic acid
BP	Bushpig (*Potamochoerus spp.*)
BSA	Bovine serum albumin
bp	Base pair(s)
cDNA	Complementary DNA
cM	Centimorgan(s) (1 million base pairs)
CMV	Cytomegalovirus
CnA	Calcineurin A
CnB	Calcineurin B
CsA	Cyclosporin A
CypA	Cyclophilin A
Da	Dalton(s)
DEPC	Diethylpyrocarbonate
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethylsulphoxide
DNA	Deoxyribonucleic acid
DNAse	Deoxyribonuclease
dNTPs	Deoxynucleotide 5'-phosphate bases
DP	Domestic pig (*Sus scrofa*)
ds	Double-stranded
EDTA	Ethylenediamine tetra-acetate
EMBL	European Molecular Biology Laboratory
EtBr	Ethidium bromide
FCS	Foetal calf serum
FL	Firefly luciferase
g	Gram(s); g-number (earth gravity - centrifugal force)
hr(s)	Hour(s)
hRL	humanised *Renilla* luciferase
HRP	Horse radish peroxidase
HS	*Homo sapiens* (human)
IκBα	Inhibitor of kappa B alpha
IPTG	Isopropylthio-beta-D-galactoside
k	Kilo (10^3)
kbp	Kilobase pair(s)
l	Litre(s)
LB	Luria broth/L-Broth
LCD	Light chain dynein
L-glut	L-glutamine
LPS	Lipopolysaccharide
Luc/luc	Luciferase

m	Milli (10^{-3})
M	Molar (moles/litre); Mega (10^{6})
n	Nano (10^{-9})
IL	Interleukin
min(s)	Minute(s)
MPS	Mononuclear phagocytic system
mRNA	Messenger RNA
N	Number of samples/replicates
NFAT	Nuclear factor of activated T cells
NFDM	Non-fat dried milk
NFκB	Nuclear factor kappa B
OD	Optical density
OIÉ	L' Office International des Épizooties
ORF	Open reading frame
p	Pico (10^{-12}); plasmid; statistical probability
P	Promoter
PA	*Phacochoerus africanus* (common warthog)
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PL	*Potamochoerus larvatus* (African bushpig)
PP	*Potamochoerus porcus* (red river hog)
Prot K	Proteinase K
PT	*Pan troglodytes* (chimpanzee)
RACE	Rapid amplification of cDNA ends
RL	*Renilla* luciferase
PMA	Phorbol-12-myristate-13-acetate
RHD	Rel homology domain
RHR	Rel homology region
RN	*Rattus norvegicus* (rat)
RNA	Ribonucleic acid
RNAse	Ribonuclease
rpm	Revolutions per minute
sec	Second(s)
SMART	Switching mechanism at 5' end of RNA transcript
SOB	SOB broth
SOC	SOC broth
ss	Single-stranded
SS	*Sus scrofa* (domestic pig or Eurasian wild boar)
SNP	Single nucleotide polymorphism
SV40	Simian virus 40
TA	Transactivation domain
TAE	Tris-acetate-EDTA
TBS	Tris-buffered saline
TBST	TBS containing 0.1% Tween 20
TE	Tris-EDTA; trypsin-EDTA
TF	Transcription factor
TK	Thymidine kinase
TM	Registered trade mark

TNF	Tumour necrosis factor
Tris	Tris(hydroxymethyl)aminomethane
UK	United Kingdom and Northern Ireland
US	United States of America
WH	Warthog (*Phacochoerus africanus*)
X-GAL	5-bromo-4-chloro-3-inolyl-beta-D-galactoside
κB	Kappa B
λ	Wavelength
μ	Micro (10^{-6})
α	Alpha
β	Beta
γ	Gamma
δ	Delta
ε	Epsilon
κ	Kappa
°C	Degrees Centigrade
1°	Primary
2°	Secondary
2n	Diploid karyotype (chromosome number)
~	Approximately

The Amino Acid Code

G - Glycine (Gly)	W - Tryptophan (Trp)
P - Proline (Pro)	H - Histidine (His)
A - Alanine (Ala)	K - Lysine (Lys)
V - Valine (Val)	R - Arginine (Arg)
L - Leucine (Leu)	Q - Glutamine (Gln)
I - Isoleucine (Ile)	N - Asparagine (Asn)
M - Methionine (Met)	E - Glutamic Acid (Glu)
C - Cysteine (Cys)	D - Aspartic Acid (Asp)
F - Phenylalanine (Phe)	S - Serine (Ser)
Y - Tyrosine (Tyr)	T - Threonine (Thr)

1 Introduction

African swine fever (ASF) poses one of the most serious threats to the global pork industry (Pan and Hess 1984; Monath 1986; Fenner *et al.* 1993). It is a highly contagious disease, which causes rapid, haemorrhagic death of domestic pigs and Eurasian wild boar (*Sus scrofa*) (Oura *et al.* 1998b; Takamatsu *et al.* 1999). No vaccine is available; the only effective means of control is by rapid identification and slaughter of infected animals (Hess 1981; Wardley *et al.* 1983; Viñuela 1985; Viñuela 1987; Wilkinson 1989). The Office International des Épizooties (OIÉ) classifies African swine fever virus (ASFV) as 'List A', placing it in the highest category of infectious animal pathogen (http://www.oie.int/) (Table 1.1). In contrast to the severe disease seen in the domestic pig, ASFV infection causes no clinical signs of disease in native African pigs, the common warthog (*Phacochoerus africanus*) and bushpig species (*Potamochoerus spp.*) (Montgomery 1921; DeTray 1963; Mansvelt 1963; Thomson *et al.* 1980; Anderson *et al.* 1998; Oura *et al.* 1998a). Such naturally occurring species-specific resistance provides a most valuable tool with which to study key molecular mechanisms integral to the pathogenesis of this disease.

At present, a number of approaches are being adopted to investigate ASF, within the broad areas of:

- Identifying ASFV genes involved in immune modulation and virulence
- Virus morphogenesis
- Host immune response to virus infection

Most current ASF research is largely virus-focussed, with relatively little work being undertaken on the host immune response to virus infection. There are however, a number of ways in which differences between the response of ASF-resistant and susceptible pig species could be investigated. One obvious approach would be to try to cross-breed these species. However, despite anecdotal evidence to suggest that this may be possible, successful breeding of warthogs or bushpigs, which have a diploid karyotype (chromosome number) of 34 (2n=34), with domestic pigs (2n=38) has

Table 1.1

The 15 diseases classified as 'List A' by the Office International des Épizooties (OIÉ), this list represents the highest category of transmissible animal pathogen. These diseases have the potential for rapid spread, irrespective of national borders, are of serious socio-economic and/or public health concern and are of major importance in the international trade of animals and animal products (http://www.oie.int/).

Disease	Infectious agent	Principle mammalian host(s)
Foot and mouth disease	Virus (Picornaviridae, *Aphthovirus*)	Ruminants and pigs
Swine vesicular disease	Virus (Picornaviridae, *Enterovirus*)	Pigs and humans
Peste des petits ruminants	Virus (Paramyxoviridae, *Morbillivirus*)	Sheep, goats and deer (Cattle and pigs asymptomatic)
Lumpy skin disease	Virus (Poxviridae, *Capripoxvirus*)	Ruminants
Bluetongue	Virus (Reoviridae, *Orbivirus*)	Ruminants
African horse sickness	Virus (Reoviridae, Orbivirus)	Equines
Classical swine fever	Virus (Flaviviridae, Pestivirus)	Pigs
Newcastle disease	Virus (Paramyxoviridae, *Rubulavirus*)	Birds (wild and domestic)
Vesicular stomatitis	Virus (Rhabdoviridae, *Vesiculovirus*)	Humans, equines, bovines, pigs and numerous small mammals.
Rinderpest	Virus (Paramyxoviridae, *Morbillivirus*)	Ruminants and pigs
Contagious bovine pleuropneumonia	Mycoplasma (*Mycoplasma mycoides mycoides*)	Domestic cattle, zebu and water buffalo
Rift Valley fever	Virus (Bunyaviridae, *Phlebovirus*)	Ruminants, monkeys and humans
Sheep pox and goat pox	Virus (Poxviridae, *Capripoxvirus*)	Sheep and goats
African swine fever	Virus (Asfarviridae, *Asfivirus*)	Pigs
Highly pathogenic avian influenza	Virus (Orthomyxoviridae, *Influenzavirus*)	Chickens and turkeys

never been demonstrated. Furthermore, even if it were possible, it is highly likely that the resulting offspring would be sterile.

Recently, a porcine immune gene/macrophage cDNA microarray has been developed by three research groups led by Dr. Stewart Lowden (University of Edinburgh), Dr. Linda Dixon (Institute for Animal Health, Pirbright) and Professor Alan Archibald (ARK Genomics, Roslin Institute). Domestic pig alveolar macrophages were isolated from lungs collected at an abattoir, cultured and infected *in vitro* with ASFV under containment conditions at IAH Pirbright. At given time points, these cells were harvested, RNA isolated and the resulting cDNA used to interrogate the microarray. These results are currently in preparation for publication and should give a valuable insight into the immune response of domestic pigs to infection. It is hoped that in the future, warthog macrophages can also be isolated and treated in the same manner. This has the potential to provide a powerful comparative analysis.

A third option would be to take a candidate gene approach. This may involve (1) identifying key host genes which are known to play an important role in ASF pathogenesis and (2) comparing these sequences between pig species of varying ASF susceptibility. Such a study would involve some of the known cellular targets of virus proteins and complement ongoing microarray findings. This latter candidate gene approach forms the basis for investigation in this thesis. Therefore, the key objectives of this thesis are as follows:

1. Comparative sequence analysis of selected host proteins targeted or mimicked by ASFV as part of its immune modulation strategy in the susceptible domestic pig (*Sus scrofa*), resistant warthog (*Phacochoerus africanus*) and phenotypically unknown babirusa (*Babyrousa babyrussa*).
2. Identification of the nature and extent of variation in the primary structure of these selected host proteins, which may be capable of modulating the different pathophysiological outcomes of ASFV infection in suid species of varying susceptibility.
3. Investigation of the role of differences in the tumour necrosis factor alpha (TNFα) promoter in mediating the proinflammatory cytokine response.

1.1 African Swine Fever: A Brief History

African swine fever (ASF) was first identified in British East Africa (Kenya) in 1909 where it was initially thought to be a variant form of hog cholera (classical swine fever) (Montgomery 1921). During this first decade, Montgomery carried out extensive research in which he described the nature of the disease, modes of transmission and first implicated native African pig species (warthogs and bushpigs) as natural reservoirs of infection. By 1926, African swine fever had also been reported in the Transvaal region of South Africa (Steyn 1928; Steyn 1932).

The first recorded outbreak of ASF outside Africa occurred in Portugal during 1957, resulting in around 450 foci of infection. This outbreak is thought to have been caused by the feeding of pigs on a farm near Lisbon airport leftover pork products from a flight originating in Angola (Manso Ribeiro *et al.* 1958). Despite an apparently successful attempt to stamp out the infection, during which 15 000 pigs were destroyed, ASF was reported again in Portugal in 1960. However, it is unclear whether this resurgence was a result of failure to completely eradicate the initial infection or represented a reintroduction of the virus. On this occasion, ASFV also spread into Spain; this resulted in a further 630 outbreaks and signalled the onset of endemic ASF in Europe. Over 120 000 pigs were slaughtered at a cost of US$9.3 million (Manso Ribeiro and Rosa Azevedo 1961; Polo Jover and Sanchez Botija 1961). Over the next twenty years, ASFV spread extensively across mainland Europe, reaching France, Madeira, Sardinia, Malta, Italy, Holland and Belgium until the virus was finally eliminated in the mid 1990s (DeTray 1963; Viñuela 1985; Monath 1986). In 1971, ASF reached the Caribbean and an outbreak on Cuba was only eradicated following the slaughter of all 430 000 pigs in the Havana Province (Hess 1981; Monath 1986). In 1978, outbreaks occurred in Brazil and the island of Hispaniola (for map see Figure 1.1). Such outbreaks posed a major threat to the pork industry in the United States (US) and prompted a report by McCauley and Sundquist at the University of Minnesota (1979, unpublished). This report estimated the economic impact of a hypothetical ASF outbreak in the US under three scenarios of increasing severity (see Table 1.2, as cited by Monath 1986).

Figure 1.1 – The Spread of African Swine Fever Virus

Map taken from Viñuela (1985), illustrating the spread of African swine fever across Europe and to the Caribbean and South America.

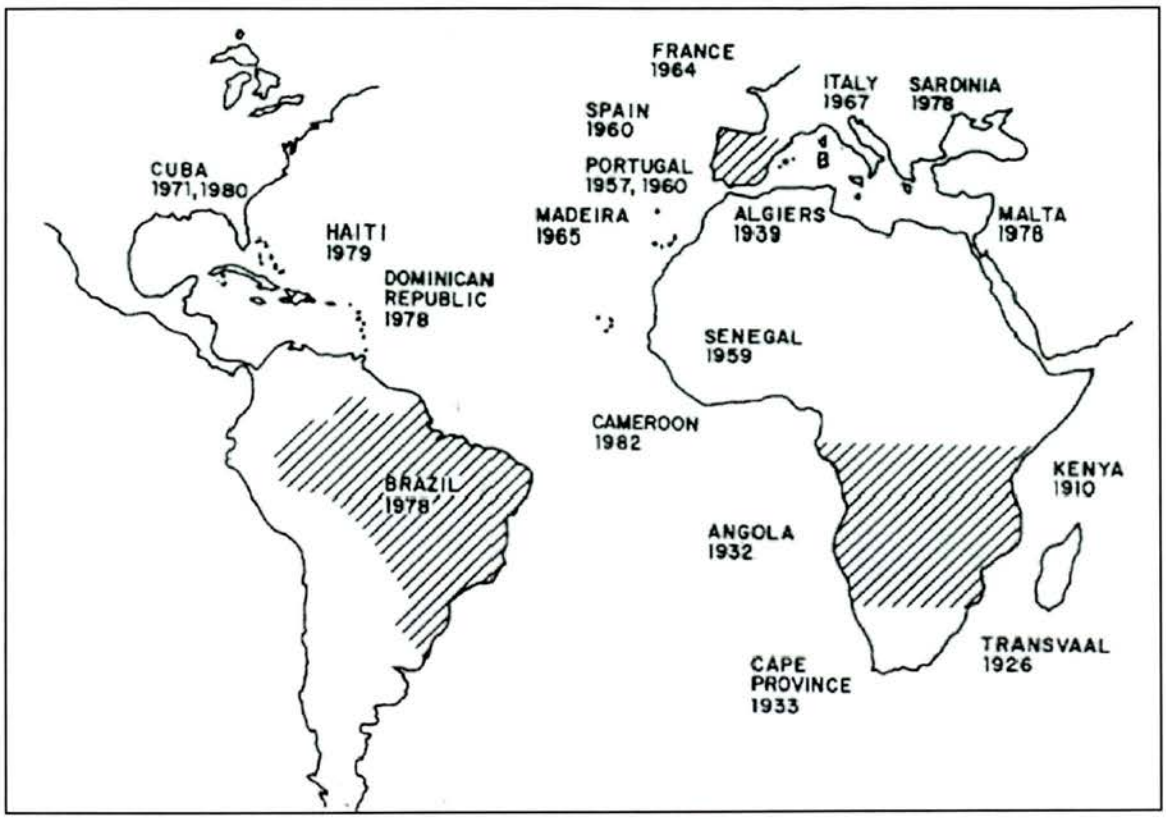

FRANCE
1964
ITALY
1967
SARDINIA
1978
SPAIN
1960
PORTUGAL
1957, 1960
MADEIRA
1965
ALGIERS
1939
MALTA
1978
CUBA
1971, 1980
HAITI
1979
DOMINICAN
REPUBLIC
1978
SENEGAL
1959
CAMEROON
1982
BRAZIL
1978
KENYA
1910
ANGOLA
1932
TRANSVAAL
1926
CAPE
PROVINCE
1933

Table 1.2

Report by McCauley and Sundquist at the University of Minnesota (1979, unpublished), in which they estimated the economic impact of a hypothetical ASF outbreak in the US under three scenarios (as cited by Monath 1986).

Scenario	US$ million
1. Successfully eradicated small outbreak:	
Direct costs (surveillance, diagnosis, slaughter, indemnification)	7.3
2. Eradication by quarantine of a larger area (e.g. state of Minnesota) over a three-year period:	
Direct costs	151.6
3. Endemic African swine fever in the US:	
Direct costs (5 years)	289.6
Direct costs (10 years)	558.6
Impact on consumer prices (5 years)	2250
Impact on consumer prices (10 years)	2300
Impact on export markets (5 years)	1500
Impact on export markets (10 years)	3000

Such was the advancing threat of ASF, that following outbreaks between 1978 and 1983, the US government ordered the destruction and repopulation of the entire pig population of Haiti and the Dominican Republic (approximately 2 million pigs at a cost of US$35 million), to prevent the risk of introduction of the virus into the US (Gibbs and Butler 1984; Monath 1986; Farmer 1994).

More recently, in 1996, an ASF outbreak in the Côte d'Ivoire triggered a two-year succession of epidemics across West Africa at a cost of US$22 million (El Hicheri *et al.* 1998; Roeder *et al.* 1999). Such outbreaks have a devastating impact on local and national economies and as such, have been the subject of major intervention by the Office International des Épizooties (OIÉ), the Food and Agriculture Organisation (FAO) of the United Nations and the World Health Organisation (WHO) (El Hicheri *et al.* 1998). ASF still represents the main threat to development of the African pig industry, where a lack of resources, adequate surveillance strategies, public awareness and government support are all serious obstacles to containing and overcoming outbreaks of ASF. Furthermore, training and provision of field staff and the information dissemination are vital. Misdiagnoses have also had catastrophic consequences and as such must be prevented. In 1998, an outbreak of ASF in Madagascar was initially mistaken for classical swine fever, the consequence of which was the slaughter of over 50% of all pig herds (Thomson 1985; Anon 2000; Anon 2001).

During the last 20 years, less virulent ASFV isolates have spread across Europe. Between 1984 and 1997, 5604 farms were infected in Spain, Portugal, France, Italy and Belgium (Monath 1986; Wilkinson 1989; Horst *et al.* 1998). These less virulent isolates have adapted to wild boar and domestic pig populations, causing subacute or chronic disease. These ASFV isolates impact significantly on production levels and are of major concern to pig producers worldwide. Furthermore, because disease is not rapidly lethal, it results in the survival and persistence of infectious hosts which shed high titres of virus.

The most recent outbreak of ASFV in mainland Europe occurred in southern Portugal during November 1999 (Sanchez-Vizcaino 1999). ASF is still endemic in

various Provinces of Sardinia, where the virus is prevalent in wild boar populations and readily transmitted to free ranging domestic pigs. So far, there have been 55 cases in Sardinia this year (as of 27th September 2004) (EU Animal Disease Notification System).

To combat the threat posed by ASF and other 'Transboundary Animal Diseases' (TADs), the FAO together with the OIÉ operates an Emergency Prevention System (EMPRES). EMPRES aims to collate and disperse information on all outbreaks of TADs, predict and monitor the situation and help implement a successful and effective control strategy (http://www.fao.org/EMPRES/default.htm).

1.2 African Swine Fever Virus

African swine fever virus (ASFV) is a large, enveloped, double-stranded DNA (dsDNA) virus. Despite having an icosahedral nucleocapsid which resembles the Iridoviruses and a genome organisation similar to the Poxviruses, ASFV is significantly different to both. It has been classified as the sole member of the Family, *Asfarviridae* (African swine fever and related viruses) and is of the Genus, *Asfivirus* (Pringle 1999; Dixon *et al.* 2000). ASFV is also the only known DNA arbovirus (being transmitted by an arthropod vector).

ASFV virions comprise an 80nm nucleoprotein core, surrounded by a protein capsid (Els and Pini 1977; Carrascosa *et al.* 1984; Andres *et al.* 1998; Rouiller *et al.* 1998; Cobbold *et al.* 2000). Capsid proteins are recruited, with the help of virus-encoded chaperones, onto the surface of the endoplasmic reticulum (ER) where they assemble into their characteristic icosahedral structure. During this process, two lipid layers derived from the ER cisternae become incorporated into the developing capsid (Cobbold *et al.* 2000; Cobbold *et al.* 2001). The nucleocapsid is surrounded by an external lipid envelope, created on budding through the plasma membrane, to form a complete virus particle ~200nm in diameter (Breese and DeBoer 1966; Moura Nunes *et al.* 1975; Els and Pini 1977; Schloer 1985; Viñuela 1987) (Figures 1.2a and 1.2b). ASFV replicates in the cytoplasm and contains a linear, covalently close-

Figure 1.2 – African Swine Fever Virus

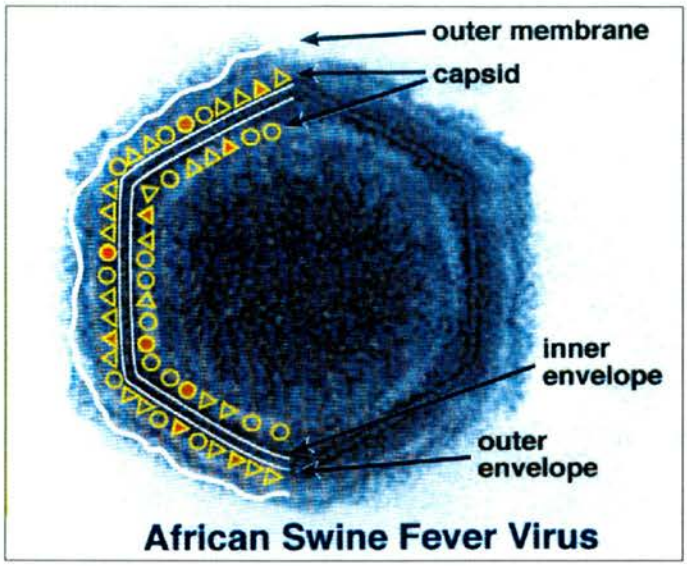
A Illustrated electron micrograph of an ASFV particle detailing the inner and outer envelopes (derived from the ER), outer membrane (derived from the cell membrane) and capsid proteins. This diagram is taken from Heath *et al.* (2001), available on the Journal of Virology website as additional material.

B Electron micrograph of ASFV particles within a perinuclear virus factory. Particles have been sectioned in different planes, leading to an apparent variation in shape.

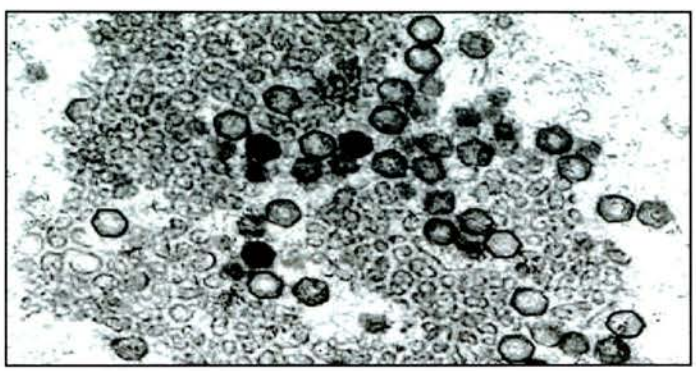
C Light micrograph of a virus factory (F) within an ASFV-infected macrophage (M). The macrophage is surrounded by haemadsorbed erythrocytes (E). The macrophage nucleus (N) can be seen displaced by the factory towards the periphery of the cell.

Diagrams B and C have been adapted from the Institute for Animal Heath Website:
http://www.iah.bbsrc.ac.uk/iah_education/ASF.html

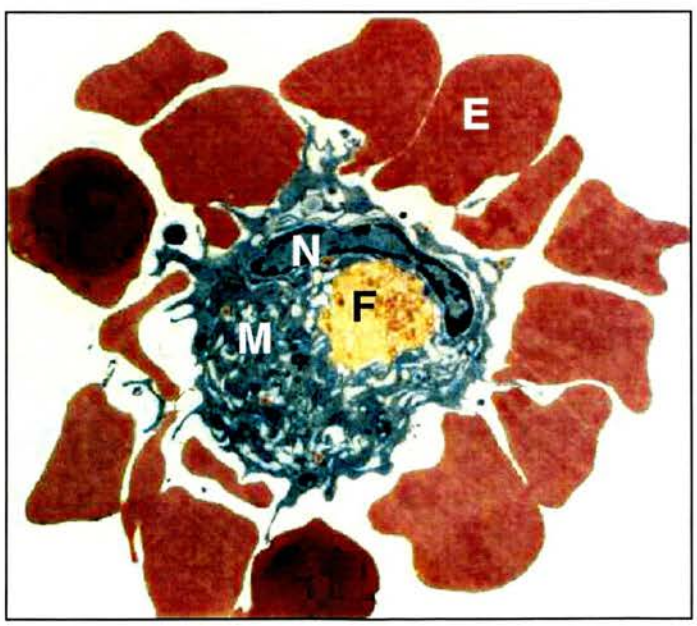
A


outer membrane
capsid
inner envelope
outer envelope
African Swine Fever Virus

B



C


E
N
F
M

ended dsDNA genome 170-190kbp in length. The end sequences are present as two 'flip-flop' forms which are inverted and complementary to each other and form hairpin loops. The viral DNA is initially replicated as head-to-head concatameric intermediates before being split into individual genomes. This bears strong resemblance to Poxvirus replication (Viñuela 1985; Dixon *et al.* 2004). The open reading frames are closely arranged and read from both strands. These encode proteins involved in infection, replication and host immune modulation (Dixon *et al.* 2000). Of the 150 genes encoded, up to 50% as yet have no known function (Tulman and Rock 2001).

The virus enters swine macrophages by receptor-mediated endocytosis and the capsid is released into the cytoplasm following fusion of the external envelope with the endocytic vesicle (Enjuanes *et al.* 1977). One receptor responsible for permissiveness to ASFV has been recently identified as CD163 (Sánchez-Torrez *et al.* 2003). Early mRNA transcription occurs in the cytoplasm using enzymes which make up some of the 30 to 40 proteins carried within the virus particle (Viñuela 1987). Virus assembly takes place within perinuclear 'virus factories', which recruit mitochondria and cellular chaperones to their surface (Moura Nunes *et al.* 1975; Heath *et al.* 2001) (Figure 1.2c). It appears that ASFV exploits the cellular machinery and uses the aggresome pathway (usually involved in the removal of misfolded cellular proteins) to concentrate viral products at these factories. The endoplasmic reticulum is also pulled into the factories, onto which the capsid proteins assemble (Carvalho *et al.* 1988; Heath *et al.* 2001). The ASFV capsid protein, p54, which is essential for productive infection, interacts with the cellular protein, light chain dynein (LCD). LCD is a cargo-specific tether which attaches the virus to cytoplasmic dynein, a minus-end-directed microtubule-associated motor complex (see Information Box 1.1). This results in retrograde transportation and localisation of virus at the microtubule organising complex (MTOC), adjacent to which the viral factories form (Rodriguez *et al.* 1996; Alonso *et al.* 2001).

Information Box 1: Cytoplasmic Dynein

Cytoplasmic dynein is a large molecular weight motor complex associated with minus-end-directed (retrograde) transport along microtubules (by contrast, axonemal dynein is involved in flagellal beating). It is also involved in chromosomal arrangement during mitosis, spindle formation, organelle maintenance (e.g. Golgi apparatus) and membranous vesicle trafficking (Holzbaur and Vallee 1994). Dynein is essential during development, with dynein-null mutant embryos failing to progress beyond the blastocyst stage (Harada *et al.* 1998). Dyneins usually consist of one to three heavy chain polypeptides and a varying number of intermediate and light chains. Each heavy chain (471-540kDa) has a globular head joined by a slender stalk to a common base. Heavy chains encode a number of distinct domains involved in dimerisation, microtubule binding, ATP hydrolysis and interaction with intermediate and light chains. Each dynein complex usually includes two intermediate chains (70-80kDa) which have a conserved C-terminal end that binds to the heavy chains and a more variable N-terminal end involved in specific cargo/organelle/light chain interaction. Approximately eight light chains (6-22kDa) are included in the complex, which are also involved in cargo/organelle-specific tethering. Of particular interest in this study is dynein light chain 8 (DLC8) (runs at 8kDa on SDS-PAGE, but is actually 10.3kDa), also known as dynein light chain 1 (DLC-1), protein inhibitor of nNOS (PIN) and light chain dynein (LCD). LCD forms as a homodimer and is highly conserved across mammals, with homologous proteins also existing in *Anthrocidaris* (sea urchin), *Aspergillus* (mould), *Drosophila* (fruit fly), *Caenorhabditis* (nematode), *Chlamydomonas* (green alga) and *Schistosoma* (trematode) (Liang *et al.* 1999; Barbar *et al.* 2001). The cargo-specific binding site of LCD exists as a deep groove within a predominantly β-sheet-derived hydrophobic pocket (Tochio *et al.* 1998). LCD is known to bind a wide variety of cargo, including the neuronal form of nitric oxide synthase (nNOS) (Jaffrey and Snyder 1996), IκBα (Crépieux *et al.* 1997), Bim (a proapoptotic Bcl-2 family member) (Puthalakath *et al.* 1999) and the p54 protein of African swine fever virus (Alonso *et al.* 2001).

1.3 Suid Classification

Within the mammalian Order Cetartiodactyla (even-toed ungulates and whales), the Suborder Suina contains two Families: Suidae (suids/pigs) and Dicotylidae/Tayassuidae (peccaries). Suina are considered to be the most primitive (basal) members of the Cetartiodactyla as they have a relatively simple two- or three-chambered stomach and retained upper incisors and canine teeth. All other Cetartiodactyl taxa have more complex stomachs (many are ruminants) and most have lost their upper incisors and canine teeth in favour of a tough dental pad (O' Leary and Geisler 1999). The suids are a diverse family, widely disseminated across Europe, Africa and Asia, whereas the three species of peccary colonise North, Central and South America. There are five genera of suid: Eurasian suids, *Sus* (domestic pigs and Eurasian wild boar), *Babyrousa* (babirusa) and Afrotropical suids, *Phacochoerus* (warthogs), *Potamochoerus* (bushpigs) and *Hylochoerus* (giant forest hogs).

1.4 ASFV Natural Hosts

The natural suid hosts for ASFV are the common warthog (*Phacochoerus africanus*) and the bushpig species (*Potamochoerus larvatus* and *Potamochoerus porcus*) (DeTray 1963; Wardley *et al.* 1983). ASFV has also been isolated on one occasion from a giant forest hog (*Hylochoerus meinertzhageni*) (Heuschele and Coggins 1965a). In these African suid species, ASFV infection causes no clinical signs of disease and can persist for long periods of time (Plowright *et al.* 1969b; Thomson *et al.* 1980; Thomson 1985; Anderson *et al.* 1998). The argasid tick (*Ornithodoros spp.*) also plays a major role in maintaining and transmitting ASFV in Africa and parts of Europe (Plowright *et al.* 1969b; Kleiboeker and Scoles 2001).

1.4.1 Warthog

The common warthog (*Phacochoerus africanus*) (2n=34) (Bosma, A A 1980; Bosma, A.A. *et al.* 1991) is disseminated widely across sub-Saharan Africa (Figure 1.3a). It has a large head with an elongated mandible, deepened maxilla and large hypsodont cheek teeth (Grubb 1993). The pelage is coarse and sparse, except for an obvious dorsal crest. Warthogs favour grazing in the open grassland and light woodland of the Savannah (Vercammen and Mason 1993). Males weigh up to 100kg with females approximately 30% lighter (Skinner and Smithers 1990). In equatorial regions warthogs breed throughout the year, but farrowing tends to coincide with the end of the dry season. The gestation period is 172 days, with an average litter size of three (ranging from one to seven). Warthogs are mainly diurnal, sleeping at night in abandoned aardvark burrows (Vercammen and Mason 1993). Warthogs are hunted as a pest species; they cause serious damage to crops and are considered competitors for grazing livestock in southern Africa (Vercammen and Mason 1993).

1.4.2 Bushpigs

The red river hog (*Potamochoerus porcus*) (2n=34) (Bosma, A.A. *et al.* 1991) inhabits the equatorial rainforests of central and western Africa (Figure 1.3b). The body colour is a bright russet orange with a distinctive white dorsal line and white tufts on the ends of elongated pinnae. The head is black with a 'mask' consisting of a white muzzle and vibrissae and white rings around the eyes. The pelage is short and soft, except for on the face where it becomes coarse and bristly (Grubb 1993; Vercammen *et al.* 1993). The red river hog is the smallest of the African suids, with males and females weighing approximately 60kg. There is a zone of intergradation in the southern Congo between the red river hog and its close relative, the African bushpig (Vercammen *et al.* 1993). There is also anecdotal evidence for interbreeding between semi-domesticated red river hog and feral domestic pig, *Sus scrofa*, in some areas of western Africa (Johnston 1905; Johnston 1906; Johnston 1908; Simoons 1953; Blom *et al.* 1990).

Figure 1.3 – The Natural Suid Hosts of ASFV in Africa and their Distribution

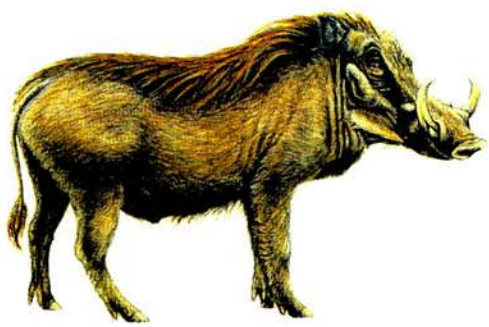
A The common warthog (*Phacochoerus africanus*) is disseminated widely across much of sub-Saharan Africa. It has a large head with an elongated mandible, deepened maxilla and large hypsodont cheek teeth. The pelage is coarse and sparse, except for an obvious dorsal crest (Skinner and Smithers 1990; Grubb 1993; Vercammen and Mason 1993).

B The red river hog (*Potamochoerus porcus*) is found in the equatorial rainforests of central and western Africa. It is the smallest of the African suids and its body colour is a bright russet orange with a distinctive white dorsal line and white tufts on the ends of elongated pinnae (Skinner and Smithers 1990; Grubb 1993; Vercammen *et al.* 1993).

C The African bushpig (*Potamochoerus larvatus*) is found across the Savannah and forests of eastern Africa and has a similar southern range to the warthog. The pelage is long and coarse over the entire body with a particularly long dorsal crest, its colouration ranges from brown to grey to black (Skinner and Smithers 1990; Grubb 1993; Vercammen *et al.* 1993).

Illustrations adapted from 'Suids of the Afrotropical Region', a poster by the IUCN Pigs and Peccaries Specialist Group.

A Common Warthog (*Phacochoerus africanus*)



B Red River Hog (*Potamochoerus porcus*)

C African Bushpig (*Potamochoerus larvatus*)

The African bushpig (*Potamochoerus larvatus*) (2n=34) (Bosma, A.A. *et al.* 1991) is the eastern and southern relative of the red river hog (Figure 1.3c). It is found throughout the highland and gallery forests of eastern Africa and has a similar southern range to the warthog. The pelage is long and coarse over the entire body with a particularly long dorsal crest, its colouration ranges from brown to grey to black. The face is never masked, but the pinnae are tufted (Grubb 1993; Vercammen *et al.* 1993). Skinner and Smithers (1990) have reported interbreeding between African bushpig and domestic pig in southern Africa.

The bushpigs (*Potamochoerus spp.*) are primarily nocturnal and do not favour an open habitat, preferring the food, water and protection offered by dense vegetation and cultivated areas adjacent to river systems (Skinner and Smithers 1990). Breeding is seasonal and coincides with the onset of the rainy season. Gestation period is 120 days and the average litter size is three (ranging from one to six), although normally only two piglets survive (Vercammen *et al.* 1993). Bushpigs, like warthogs, are the cause of much crop damage. Expansion of agriculture in place of former forest and/or scrub has led to an increase in their numbers. Efforts by farmers to eliminate bushpigs have proved unsuccessful. Hence neither species is considered threatened (Vercammen *et al.* 1993).

1.4.3 *Ornithodoros* Tick

The soft-bodied argasid tick, *Ornithodoros moubata*, is a vector and reservoir host of African swine fever virus (Figure 1.4) (Sánchez Botija 1963; Heuschele and Coggins 1965b; Plowright *et al.* 1974). In Spain and Portugal, a related tick species, *Ornithodoros marocanus* (formerly *O. erraticus*), has been also identified as a vector for ASFV (Sánchez Botija 1963; Plowright *et al.* 1969a; Kleiboeker and Scoles 2001).

In addition to becoming infected with ASFV during a blood feed, female *Ornithodoros* ticks may become infected by copulation with an infected male

Figure 1.4 – The *Ornithodoros* Tick

The soft bodied tick, *Ornithodoros spp*, acts as a vector and reservoir host of African swine fever virus.

A Dorsal view

B Ventral view

Images taken from the Wilderness Medical Society website:
http://www.wms.org/Templates/Ticks%20that%20transmit%20diseases.htm

A

B

(Plowright *et al.* 1974). Transovarial (from the body cavity into the ovary) and transstadial (when the virus survives successive moults) transmission of ASFV also occurs (Plowright *et al.* 1969a; Plowright *et al.* 1970; Greig 1972; Rennie *et al.* 2001). Together with venereal transmission, these modes of transmission represent essential mechanisms for sustaining infection in the tick population in the absence of viraemic suid hosts (Plowright *et al.* 1974; Rennie *et al.* 2001).

In ticks, ASFV particles are first seen in phagocytic cells of the midgut epithelium where it replicates before slowly crossing the basal lamina and entering the haemocoel (15 to 21 days post infection). Here it infects and undergoes secondary replication in connective tissue, haemocytes (type I and II), coxal gland, salivary gland and reproductive tissue. High titres of virus are seen in coxal and salivary glands and are maintained at this level throughout infection. Such high levels may be necessary for transmission, because *Ornithodoros* ticks feed rapidly (within 1 hour), unlike ixodid (hard) ticks, which may feed for a number of days (Kleiboeker *et al.* 1998).

1.4.4 Other Potential Hosts/Vectors

In addition to the Afrotropical suids and *Ornithodoros moubata* ticks, many other species have been investigated as potential hosts of ASFV.

All ixodid (hard) ticks studied have proved incapable of maintaining and transmitting ASFV (Sánchez Botija 1963; Kovalenko and Sidorov 1973; Plowright 1977; Groocock *et al.* 1980; Plowright *et al.* 1994), whereas all North American and Caribbean *Ornithodoros* tick species tested have been successfully infected as potential vectors of ASFV (Groocock *et al.* 1980; Gibbs and Butler 1984; Hess *et al.* 1987; Kleiboeker and Scoles 2001). One study demonstrated that the pig louse (*Haematopinus suis*) can transmit ASFV between domestic pigs (Sanchez Botija and Badiola 1966), although this result has not been reproduced (Montgomery 1921; Heuschele and Coggins 1965b; Kovalenko and Sidorov 1973). The haematophagous larvae of the Congo floor maggot (*Auchmeromyia luteola*) are known to be involved

in Trypanosome transmission between warthogs, but their role in ASF epidemiology remains unknown (Boreham and Geigy 1976). In addition, the common stable fly (*Stomoxys calcitrans*) has been shown to act as a mechanical vector of ASFV between domestic pigs (Mellor *et al.* 1987).

Other mammals have also been implicated and subsequently ruled out as potential reservoirs of ASFV by experimental infection. These include the common hippopotamus (*Hippopotamus amphibius*), hyena (*Crocuta spp.*) and porcupine (*Histrix spp.* and *Atherurus spp.*) (Cox 1963; Stone and Heuschele 1965). There are no published reports of the fifth genus of suid (*Babyrousa*) having ever been challenged with ASFV.

1.4.5 Babirusa

The endangered babirusa (*Babyrousa babyrussa*) (2n=38) is primarily located on the island of Sulawesi in the Indonesian archipelago (Bosma, A A 1980; Bosma, A A and de Haan 1981; Macdonald 1993; Bosma, A A *et al.* 1996) (Figure 1.5). It is considered to be the most ancient species of extant suid, with molecular evidence indicating that it has no common ancestor with the domestic pig more recently than 10-19 million years ago (Randi *et al.* 1996). Indeed, interpretation of the fossil record and morphological-based analyses have suggested an even earlier date of up to 40 million years ago (Thenius 1970; Groves and Grubb 1993; O' Leary and Geisler 1999). In this study, the babirusa is considered an outgroup suid species of unknown ASFV-response phenotype.

1.5 ASFV Infection of the Domestic Pig (*Sus scrofa*)

1.5.1 Clinical Manifestations

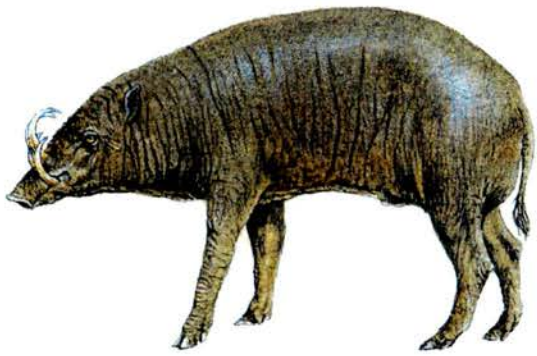
As previously described, ASFV isolates differ in virulence. In the domestic pig, ASFV causes disease ranging from peracute to chronic forms (Hess 1981). In

Figure 1.5 – The Babirusa

The babirusa (*Babyrousa babyrussa*) is primarily located on the island of Sulawesi in the Indonesian archipelago. A distinctive feature of the male babirusa is the upper canines with grow vertically from the maxillary alveoli and curve caudally and medially in front of the face (Macdonald 1993).

Illustration adapted from 'Suids of South-East Asia',
a poster by the IUCN Pigs and Peccaries Specialist Group.

Babirusa (*Babyrousa babyrussa*)



peracute forms, animals usually die suddenly without any prior sign of illness and minimal gross lesions. Widespread haemorrhage is the most prominent post-mortem finding (De Kock *et al.* 1940; Maurer *et al.* 1958; Moulton and Coggins 1968; Hess 1981; Edwards *et al.* 1984). Pigs suffering from **acute** ASF are pyrexic for one to two days (40-42°C), become anorexic and develop erythemic or cyanotic regions (snout, ears, fetlocks, ventral abdomen and hindquarters), which subsequently become haemorrhagic (Figures 1.6a and 1.6b). Some animals also display respiratory distress, vomiting, polydypsia and occasionally bloodstained faeces (Maurer *et al.* 1958; Colgrove *et al.* 1969) (Figures 1.6c and 1.6d). Death usually occurs within seven to ten days. Post-mortem findings include severe damage to the vascular system (characterised by degeneration of vascular endothelium), splenomegaly and widespread haemorrhages in most organ systems, particularly the visceral lymph nodes. This is coupled with extensive oedema and ascites in the pleural and peritoneal cavities and pericardial sac. In some cases, bleeding from body orifices is also seen (Steyn 1928; De Kock *et al.* 1940; Manso Ribeiro *et al.* 1958; Mebus *et al.* 1983; Neser *et al.* 1986) (Figure 1.7). **Subacute** forms of ASF cause persistent pyrexia or fluctuating body temperature for up to 20 days. Some individuals display signs similar to those with the acute form of the disease. Mortality rate is variable (Hess 1981) and abortion is seen in 90% of infected pregnant animals (McDaniel 1978; Wilkinson *et al.* 1980; Wardley *et al.* 1983). **Chronic** ASF is highly variable and may persist for many months. Symptoms may include pneumonia, stunting, emaciation, arthritis and skin ulcers (Hess 1981). Hypergammaglobinaemia is frequently observed and may be due to widespread (non-specific) activation of B lymphocytes (Takamatsu *et al.* 1999).

1.5.2 Cellular Pathophysiology

Despite the significant body of research, the precise mechanisms of ASFV pathophysiology in domestic pigs are still not completely understood. This is largely due to the complex nature of the disease, coupled with the inherent difficulties associated with working with a list A pathogen. This section summarises current evidence on the ASF disease process.

Figure 1.6 – African Swine Fever: Clinical Signs

African swine fever-associated pathology in the live animal.

A Characteristic erythemic (red) ventral abdomen and distal limbs seen during ASFV-infection. This region subsequently becomes cyanotic (blue) and eventually haemorrhagic.

B Multiple cutaneous haemorrhages (purpura) and congestion of the skin.

C Severe congestion and haemorrhage of the ocular mucous membranes.

D Bloody diarrhoea caused by ASF

Photographs taken from the Food and Agricultural Organisation (FAO) Animal Health Manual 'Recognizing African Swine Fever - A field manual' (Anon 2001) and the FAO Emergency Prevention System (EMPRES) online photolibrary: (http://www.fao.org/WAICENT/FAOINFO/AGRICULT/AGA/AGAH/EMPRES/GEMP/resources/resources.html)

A

B

C

D

Figure 1.7 – African Swine Fever: Post-mortem Findings

African swine fever-associated post-mortem pathology.

A Numerous petechical haemorrhages (petechiae) across the surface of the kidney.

B Diffuse haemorrhage in the pelvis of the kidney.

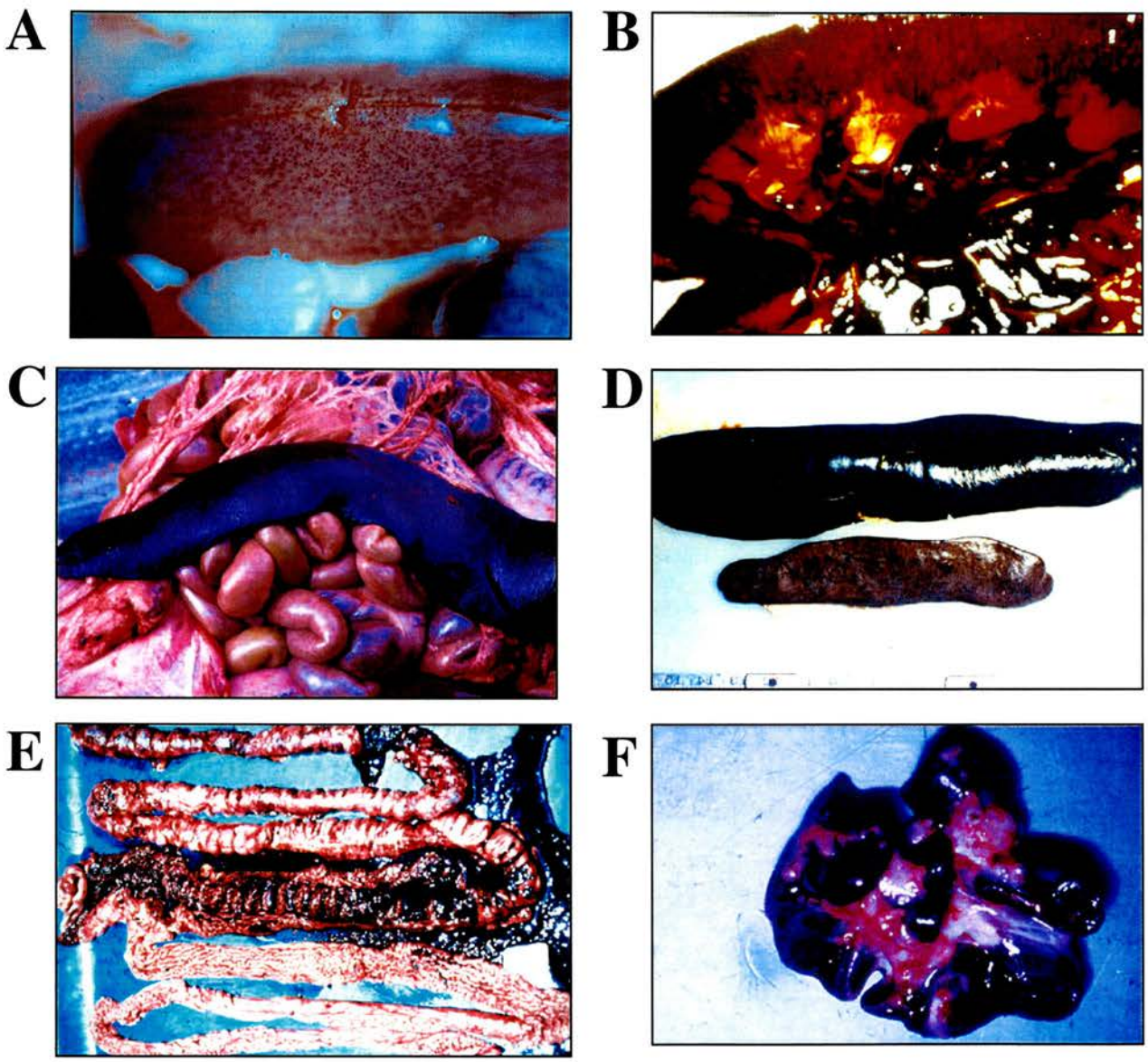
C Splenomegaly (enlarged spleen) and haemorrhagic intestines.

D Splenomegaly seen in ASFV-infected pigs (above) compared to normal (below).

E Haemorrhagic intestines.

F Enlarged, haemorrhagic mesenteric lymph node.

Photographs taken from the Food and Agricultural Organisation (FAO) Animal Health Manual 'Recognizing African Swine Fever - A field manual' (Anon 2001) and the FAO Emergency Prevention System (EMPRES) online photolibrary: (http://www.fao.org/WAICENT/FAOINFO/AGRICULT/AGA/AGAH/EMPRES/GEMP/resources/resources.html)


A
B
C
D
E
F

1.5.2.1 Cellular Targets

The primary site of viral replication is within cells of the mononuclear phagocytic system (MPS), which include monocytes and macrophages (Colgrove 1968; Moulton and Coggins 1968; Wardley and Wilkinson 1977; Casal *et al.* 1984; Pan 1987; Fernández *et al.* 1992a; Fernández *et al.* 1992b; Carrilllo *et al.* 1994). Virulent ASFV isolates initially infect extrafollicular macrophages and cells within the epithelium of the tonsillar crypts before proceeding to infect macrophages of the spleen and other lymphoid tissues (Colgrove *et al.* 1969; Fernández *et al.* 1992c; Oura *et al.* 1998b; Carrasco *et al.* 2002). However, during the later stages of infection, replication has also been described in a range of other cells including vascular endothelial cells (see Table 1.3). This infection of non-MPS cells may be associated with the release of cytokines by infected macrophages (Gómez-Villamandos *et al.* 1995) as described in Section 1.5.2.2.

1.5.2.2 Haemorrhagic Fever

Haemorrhagic fever is the most prominent feature of ASFV infection in the domestic pig. Initially, haemorrhage was thought to be due to vascular damage as a result of infection and destruction of endothelial cells (Maurer *et al.* 1958; Colgrove *et al.* 1969; Wilkinson and Wardley 1978). However, no significant infection of endothelial cells is seen *in vivo* until the later stages of disease (Mebus *et al.* 1983; Oura *et al.* 1998b), indicating that the pathogenic effects mediated by ASFV are not a direct result of infection of endothelial cells (Gómez del Moral *et al.* 1999).

There is evidence that the haemorrhagic pathology may be due to inappropriate cytokine release by infected macrophages, resulting in an 'autoinflammatory' condition similar to septic shock (Gómez-Villamandos *et al.* 1996; Gómez del Moral *et al.* 1999; Carrasco *et al.* 2002; Salguero *et al.* 2002; Gil *et al.* 2003). In this model, the virus-induced systemic proinflammatory cytokine release results in widespread activation of the vascular endothelium. During this process, selectins and cell adhesion molecules (for example E-selectin, VCAM-1) are upregulated on the

Table 1.3

Examples of the range of non-MPS cell types within which ASFV replication has been described during the later stages of infection.

Non-MPS cell type	Reference(s)
Vascular endothelial	(Moulton and Coggins 1968; Colgrove *et al.* 1969; Wilkinson and Wardley 1978; Pan 1987; Fernández *et al.* 1992b)
Megakaryocytes	(Edwards *et al.* 1985)
Platelets	(Edwards and Dodds 1985; Neser *et al.* 1986)
Neutrophils	(Carrasco *et al.* 1996)
Smooth muscle	(Gómez-Villamandos *et al.* 1995a)
Hepatocytes	(Sierra, AM *et al.* 1987; Fernández *et al.* 1992a)
Glomerular mesangial cells	(Sierra, MA *et al.* 1989; Gómez-Villamandos *et al.* 1995b)
Epithelial cells of renal collecting ducts	(Sierra, MA *et al.* 1989; Gómez-Villamandos *et al.* 1995b)
Kupffer cells	(Sierra, AM *et al.* 1987)

.

luminal surface of activated endothelial cells, facilitating the migration of neutrophils into the tissues. In addition to degranulation and phagocytosis, neutrophils in turn release a variety of cytokines, which activate and recruit more leukocytes into the tissues, exacerbating the magnitude of the inflammatory response. Furthermore, cytokine release and endothelial cell activation trigger the activation of platelets and the clotting cascade. Platelet counts suddenly decrease with the appearance of large numbers of ASFV-infected, activated macrophages (Edwards *et al.* 1984; Edwards and Dodds 1985; Gómez-Villamandos *et al.* 1995). This may explain the marked thrombocytopaenia and disseminated intravascular coagulation (DIC) seen during ASFV infection (Edwards *et al.* 1985; Neser *et al.* 1986). DIC causes massive consumption of clotting factors, loss of plasma volume and failure of the clotting response. In addition to triggering production of acute phase proteins (APPs) by hepatocytes, interleukin (IL) 1, IL6 and tumour necrosis factor alpha (TNFα) also generate fever by acting on the thermoregulatory centre of the hypothalamus.

It would appear that ASFV-induced, prolonged systemic cytokine production and subsequent endothelial dysfunction, results in haemorrhage and widespread collection of plasma in tissues and capillaries. The resulting hypovolaemic shock causes collapse of major blood vessels and ultimately multiple organ failure.

1.5.2.3 Lymphocyte Apoptosis (Lymphopaenia)

In addition to haemorrhagic fever, widespread 'bystander apoptosis' of uninfected lymphocytes is seen during infection of the domestic pig (Gómez-Villamandos *et al.* 1995; Ramiro-Ibáñez *et al.* 1996; Oura *et al.* 1998b). This is similar to pathology caused by a range of other viral haemorrhagic fevers including classical swine fever (Sato *et al.* 2000) and Ebola (Geisbert *et al.* 2000). Paradoxically, ASFV infection initially triggers non-specific activation and proliferation of lymphocytes, which are resistant to ASFV infection, before widespread apoptosis subsequently occurs (Takamatsu *et al.* 1999). By day seven post infection, the white pulp of the spleen is completely destroyed and red pulp is almost entirely depleted of cells. Intense apoptosis also occurs within lymph nodes, seen as broad bands of apoptotic T cells.

During the later stages of infection, this apoptosis extends to include B cells in follicular areas. The degree of apoptosis correlates strongly with the number of infected macrophages present (Ramiro-Ibáñez *et al.* 1996; Oura *et al.* 1998b) and once again may be related to proinflammatory cytokine production (Lenardo *et al.* 1999). In support of this theory, the TNFα-containing supernatant from ASFV-infected cell cultures has been shown to induce apoptosis in uninfected lymphocytes (Gómez del Moral *et al.* 1999). Furthermore, hypergammaglobinaemia is observed during chronic infection, a phenomenon which is often attributed to cytokine expression, rather than viral antigens (Takamatsu *et al.* 1999). A number of additional mechanisms have been proposed to explain the systematic apoptosis of lymphocytes seen during ASFV infection. These range from direct virus-mediated interference with the apoptosis pathway (Ramiro-Ibáñez *et al.* 1996), to the suggestion that an absence of T cells, due to intense apoptosis, in turn deprives proliferating B cells of the CD154 signal (CD40 ligand) they require for survival. This process may result in later B cell apoptosis (Takamatsu *et al.* 1999).

Thus it appears that events which influence the balance between initial lymphocyte proliferation and subsequent apoptosis may represent critical factors in determining the pathological outcome of ASFV infection.

1.6 Cytokine Response to ASFV Infection

It is apparent that the key pathophysiological events described above, including haemorrhagic fever, thrombocytopaenia, vascular endothelial activation, lymphocyte apoptosis, hypergammaglobinaemia, neutrophilia and infection of non-MPS cells all coincide with the appearance of large numbers of infected, activated macrophages.

1.6.1 The Macrophage and Cytokine Production

Macrophages represent one of the most important cells of the immune system. They are involved in 'front-line' phagocytic defence, immunosurveillance, antigen presentation and have the ability to rapidly produce an array of potent inflammatory

mediators in response to a wide variety of pathogens. This powerful system is usually maintained under strict control. However, when macrophages themselves become infected with, or are exposed to, ASFV they release large amounts of proinflammatory cytokines, including TNFα, IL1 and IL6. As previously described, this massive cytokine release can have a catastrophic local and systemic impact, including widespread cellular activation, leukocyte recruitment, uncontrolled apoptosis and vascular dysfunction (Oura *et al.* 1998b; Gómez del Moral *et al.* 1999; Lenardo *et al.* 1999; Carrasco *et al.* 2002; Salguero *et al.* 2002).

It remains unclear whether the massive cytokine release by macrophages seen during ASF is due to virus infection of macrophages, which directly alters their function and behaviour, or is due to an excessive response by macrophages to the presence of virus, which is poorly directed and effectively results in an autoinflammatory disease.

1.6.2 Tumour Necrosis Factor Alpha (TNFα)

TNFα is one of the principle proinflammatory cytokines to be upregulated during ASFV infection and has previously been associated with pathology caused by a wide range of other viral haemorrhagic fevers, infectious and autoimmune diseases (Table 1.4). Tumour necrosis factor (TNF) was originally named after its ability to cause some infected tumours to become necrotic and regress. It was initially thought that TNF was released by invading bacteria, until it was discovered that bacterial-derived lipopolysaccharide (LPS) was in fact stimulating release of TNF by the host (Carswell *et al.* 1975; Aggarwal *et al.* 1985; Old 1985).

TNFα is a 51kDa homotrimeric, pluripotent cytokine which performs a broad range of functions from triggering apoptosis to promoting cell survival and proliferation. TNFα is expressed either as a membrane-bound form or as a soluble cytokine when cleaved by TNF alpha converting enzyme (TACE). TNFα is a member of the TNF superfamily which comprises more than 17 membrane-bound and/or secreted ligands, which bind to members of a corresponding superfamily of TNF receptors.

Table 1.4

Examples of the range of other viral haemorrhagic fevers, infectious and autoimmune diseases which have be associated with TNFα production.

Disease	Reference(s)
Argentine haemorrhagic fever	(Marta *et al.* 1999)
Ebola haemorrhagic fever	(Villinger *et al.* 1999; Geisbert *et al.* 2000; Baize *et al.* 2002)
Marburg haemorrhagic fever	(Feldmann *et al.* 1996)
Dengue haemorrhagic fever	(Hober *et al.* 1993; Marianneau *et al.* 1998; Rothman and Ennis 1999; Azeredo *et al.* 2001; Kurane and Takasaki 2001)
Pichinde virus infection	(Aronson *et al.* 1995)
Hanatvirus pulmonary syndrome	(Mori *et al.* 1999)
Classical swine fever	(Sato *et al.* 2000)
Cerebral malaria	(Knight *et al.* 1999)
Endotoxic/septic shock	(Goldman *et al.* 2001; Pfeffer 2003)
Trypanosomiasis	(Okomo-Assoumou *et al.* 1995; Magez *et al.* 2004)
Multiple sclerosis	(Kassiotis and Kollias 2001)
Rheumatoid arthritis	(Kaijzel *et al.* 2001)
Sarcoidosis	(Baughman *et al.* 2003)
Systemic lupus erythematosus (SLE)	(Wilson *et al.* 1995; Kollias and Kontoyiannis 2002)

To date, around 40 different combinations TNF ligands and TNF receptors have been identified, due to extensive ligand/receptor cross-reactivity (MacEwan 2002; Aggarwal 2003; Wajant *et al.* 2003; Ware 2003).

1.6.3 TNF Receptor Signalling

Approximately 29 TNF receptor superfamily members have been identified in humans. Similar to TNF ligands, TNF receptors form homotrimers which may also be proteolytically cleaved and shed from the cell surface. Receptor shedding may represent a protective mechanism to guard against over-induction by TNF ligands, both by dissociating from their intracellular signalling pathways and by 'mopping up' free TNF ligand (Magez *et al.* 2004).

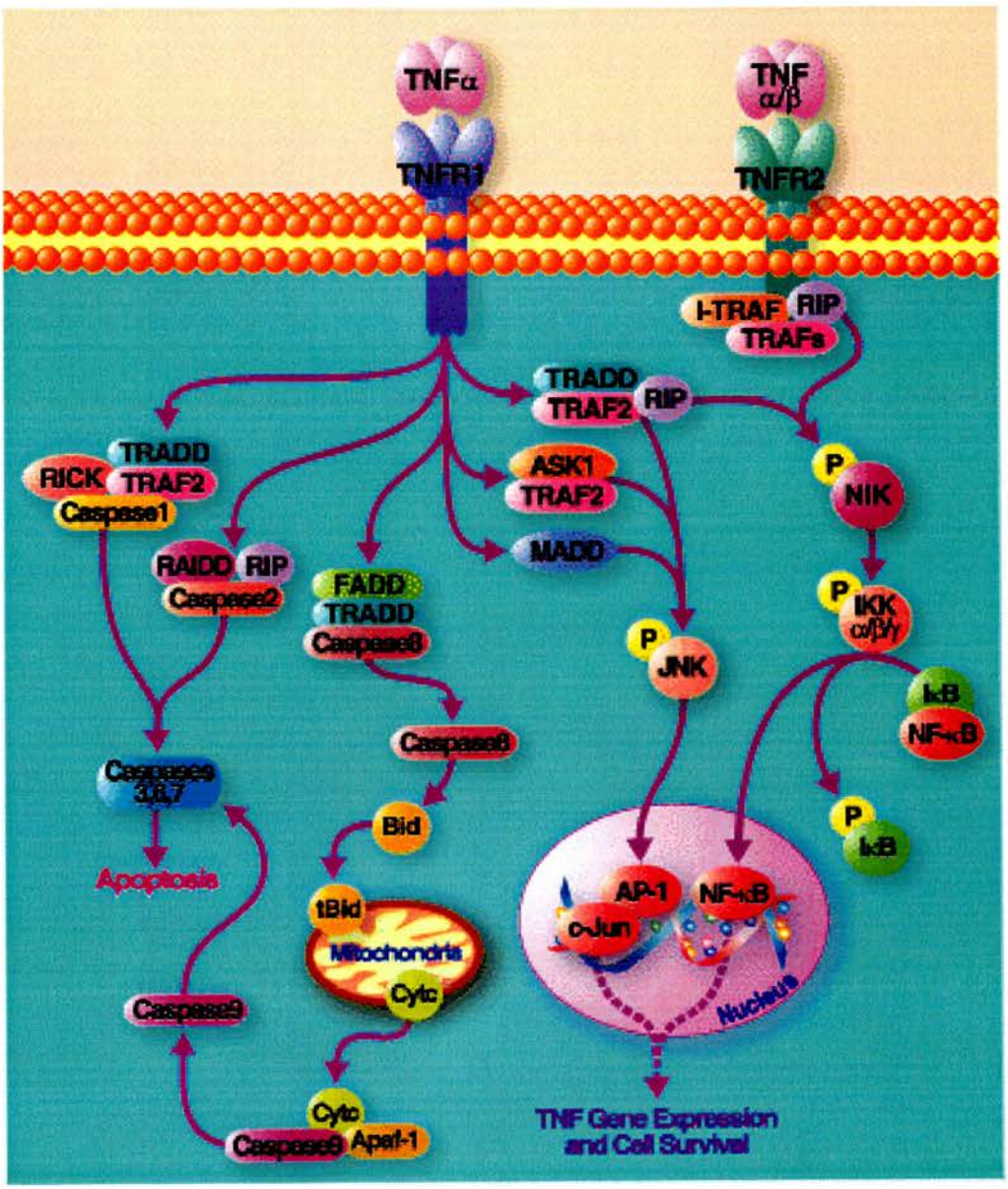
TNF receptors fall into three major categories, based on their cytoplasmic domains (Locksley *et al.* 2001). The first group includes TNF receptor 1 (TNFR1) and Fas, which contain a death domain (DD) in their cytoplasmic tail. Activation of this group can result in recruitment of intracellular death domain adaptors such as TNFR-associated death domain (TRADD) and Fas-associated death domain (FADD). These adaptors subsequently activate the caspase cascade and induce type I apoptosis. The second group of TNF receptors include TNF receptor 2 (TNFR2) and CD40. These contain motifs in their cytoplasmic domains, which interact with TNF receptor associated factors (TRAFs). When group 2 receptors are activated, their TRAF-interacting motifs (TIMs) recruit TRAFs. These in turn activate multiple signal transduction pathways, including nuclear factor kappa B (NFκB), c-Jun NH_2-terminal kinase (JNK), p38, extracellular signal-related kinase (ERK) and phosphoinositide-3-kinase (PI3K). These pathways regulate a wide variety of cellular processes from cell proliferation and differentiation to immune response and apoptosis (Figure 1.8). The third group of TNF receptors do not have active signalling domains. However, they compete with group 1 and 2 receptors for their corresponding ligands (Dempsey *et al.* 2003). TNFα acts primarily on TNFR1 and TNFR2, which are also referred to in the literature as TNFR1A/p55/CD120a and TNFR1B/p75/CD120b respectively.

Figure 1.8 – Tumour Necrosis Factor Alpha (TNFα) Signalling

Diagrammatic representation of the signaling pathways mediated by TNFα, through TNF receptor (TNFR)-1 and TNFR2. TNFR1 may initiate either cell survival and proliferative mechanisms (antiapoptotic) or apoptotic pathways. TNFR2 activates only cell survival pathways and may act to mediate the apoptotic signals from TNFR1.

Abbreviation	Description
TNFα	Tumour necrosis factor alpha
TNFR	Tumour necrosis factor receptor
RIP	Receptor-interacting protein
TRAF	TNFR-associated factor
I-TRAF	TRAF-interacting factor
TRADD	TNFR-associated death domain
RICK	Also called Rip2 (a serine/threonine kinase containing a caspase recruitment domain (CARD))
RAIDD	Casapase-2 adapter protein with a death domain
FADD	Fas-associated death domain
MADD	MAPK-activating death domain-containing protein
MAPK	Mitogen-activated protein kinase (serine/threonine kinases)
Bid	Bcl-2 family member (proapoptotic)
tBid	Truncated form of Bid
Cytc	Cytochrome c
Apaf-1	Apoptotic protease-activating factor 1
P	Phosphorylated protein/complex
NIK	Nuclear Factor kappa B-inducing kinase
JNK	c-Jun NH_2-terminal kinase (a MAPK)
IKK	I kappa B kinase complex (contains α, β and γ subunits)
IκB	Inhibitor of (nuclear factor) kappa B
NF-κB	Nuclear Factor kappa B (transcription factor)
AP-1	Activator Protein-1 (transcription factor complex)
c-Jun	c-Jun (major component of AP-1)

Figure adapted from 'TNF Signaling', an online resource from The Biotech Journal: http://www.biotechjournal.com/


TNFα
TNF α/β
TNFR1
TNFR2
I-TRAF
RIP
TRAFs
TRADD
TRAF2
RIP
TRADD
RICK
TRAF2
Caspase1
ASK1
TRAF2
P
NIK
RAIDD
RIP
Caspase2
MADD
FADD
TRADD
Caspase8
P
JNK
P
IKK
α/β/γ
IκB
NF-κB
Caspase8
Caspase
3,6,7
Bid
P
IκB
Apoptosis
tBid
AP-1
NF-κB
c-Jun
Mitochondria
Cytc
Nucleus
Caspase9
Cytc
Caspase9
Apaf-1
TNF Gene Expression
and Cell Survival

1.6.4 TNFα Promoter

The gene encoding TNFα is located on chromosome seven in the domestic pig, adjacent to the major histocompatibility complex (MHC) locus. TNFα expression is regulated at several levels, including transcription, mRNA stability, translation and post-translational modification. Transcription of TNFα is primarily driven by its *cis*-acting promoter, which lies in the ~1.6kbp region, separating the TNFα coding sequence from the upstream gene encoding TNFβ (Kuhnert *et al.* 1991; Solinas *et al.* 1992). The TNFα promoter is rich in transcription factor binding domains, including those for NFκB, NFAT, activator proteins, stimulating proteins, serum response factors, interferon regulatory factors, CAAT enhancer-binding protein (C/EBP), tumour suppressor p53 and signal transducers and activators of transcription (STATs).

In vitro expression studies comparing the human and baboon TNFα promoter indicate that polymorphisms can significantly alter expression levels and sensitivity in response to infection (Haudek *et al.* 1998). In addition, polymorphisms between geographically distinct human populations have been related to the response to and outcome of diseases such as cerebral malaria (Knight *et al.* 1999). It is therefore possible that variation in the TNFα promoters between ASFV-susceptible and ASFV-resistant suid species may determine their TNFα expression profile in response to ASFV infection. This may cause them to be more or less susceptible to producing a massive 'cytokine storm' in response to infection.

Although many single nucleotide polymorphisms (SNPs) have been identified in the human TNFα gene, recently there has been much interest in SNPs at positions -238, -308 and -376 (with respect to the transcription start point). These polymorphisms have been associated with susceptibility to disease, progression of disease and responsiveness to different forms of therapy (Wilson *et al.* 1995; Knight and Kwiatkowski 1999; Knight *et al.* 1999; Hajeer and Hutchinson 2000; Hajeer and Hutchinson 2001; Kaijzel *et al.* 2001; Neben *et al.* 2002).

1.7 ASFV Infection in the Natural Hosts

The natural suid hosts of ASFV (warthogs and bushpigs) are readily infected with ASFV. Unlike the domestic pig however, they develop a subclinical form of the disease from which they rapidly recover (Montgomery 1921; DeTray 1963; Mansvelt 1963; Thomson *et al.* 1980; Anderson *et al.* 1998; Oura *et al.* 1998a).

ASFV infection of warthogs relies on the *Ornithodoros* tick vector or consumption of an infected carcass (Cox 1963; Plowright *et al.* 1969b). Neither direct, horizontal nor vertical transmission has been demonstrated between warthogs (Plowright *et al.* 1969b). Adult warthogs are often found to be seropositive to ASFV (6-14 months post-infection), but the virus is largely restricted to lymphoid tissue and levels in the blood are not high enough to infect feeding ticks (Heuschele and Coggins 1965b; Plowright *et al.* 1969b; Thomson *et al.* 1980; Anderson *et al.* 1998). *Ornithodoros moubata* ticks are largely nidicolous (nest/burrow-dwelling) and several thousand may inhabit a single burrow in which warthogs sleep (Plowright *et al.* 1969b; Plowright *et al.* 1969a; Kleiboeker and Scoles 2001). Neonate warthogs are restricted to these burrows and rapidly become infected with ASFV by feeding ticks. The resulting viraemia causes neonate warthogs to be infectious to ticks for up to 3 weeks (days 37-59 post-infection) and represents a principle mechanism for cyclical ASFV infection of both ticks and warthogs (Thomson 1985; Anderson *et al.* 1998; Kleiboeker and Scoles 2001). Direct contact transmission from warthogs to domestic pigs does not occur (Montgomery 1921; DeTray 1963; Heuschele and Coggins 1969; Thomson 1985). Interestingly however, African bushpigs have been shown to transmit a virulent ASFV isolate directly to domestic pigs kept in an adjacent pen (Anderson *et al.* 1998).

In bushpigs, ASFV can only be detected for around eight months post-infection in the lymphatic tissue (compared with up to 14 months in the warthog). However, during this time the level of viraemia is high enough to infect feeding ticks for seven weeks (days 30 to 80 post-infection) (Anderson *et al.* 1998). It has been suggested that 'controlled' apoptosis of uninfected B lymphocytes (T cells are unaffected) in the bushpig may act to suppress the immune response and decrease antibody

production for long enough to allow viral persistence and the opportunity for transmission (Oura *et al.* 1998a).

Although warthogs are more widespread and numerous than bushpigs (Skinner and Smithers 1990) and are more frequently infected with ASFV (Mansvelt 1963), they inhabit the open savannah and rarely come into contact with domestic pigs (Skinner and Smithers 1990; Anderson *et al.* 1998). Furthermore, warthogs rely almost entirely on the tick vector for transmission and after a brief initial viraemia, the virus is restricted to lymphatic tissue and is inaccessible to feeding ticks (Heuschele and Coggins 1965b; Plowright *et al.* 1969b; Thomson *et al.* 1980; Thomson 1985; Kleiboeker and Scoles 2001). In contrast, the nocturnal activity of bushpigs and their preference for areas of cultivation bring them more frequently into contact with areas of human habitation and domestic pigs (although their presence often goes unnoticed). This suggests that contrary to earlier beliefs, bushpigs may be primarily responsible for transmission of ASFV to domestic pigs.

Once in the domestic pig population, ASFV is highly contagious and transmitted by direct contact, by the tick vector and via infected meat. ASFV is excreted in urine and faeces, but much higher titres of virus are found in the nasal and pharyngeal secretions (Montgomery 1921; Greig and Plowright 1970). Fomites in the form of contaminated buildings, animal bedding, vehicles, feeding/drinking troughs, clothing and veterinary instruments (e.g. hypodermic needles) have all been identified as being capable of retaining and transmitting ASFV (Anon 2001).

1.8 Immune Mechanisms Associated with Viral Infection

Viruses are obligate intracellular parasites which must enter a host cell and hijack its nucleic acid and protein synthesis machinery in order to replicate, package and disseminate their genome. To protect themselves from such an assault, many organisms have developed an immune system in order to detect and destroy invading pathogens. In mammals, this system is highly complex and comprises both innate and adaptive components. Consequently, in order to prevent detection and

elimination, many viruses have in turn evolved elaborate mechanisms to evade or modulate the host immune response (Tizard 1996; Krajcsi and Wold 1998; Haig 2001).

DNA viruses in particular are renowned for their ability to encode homologues of cellular genes which modulate the host immune response. For example, poxviruses produce a TNFR2 homologue which 'mops up' TNFα, preventing it from having any biological action (Benedict and Ware 2001). Vaccinia virus, also a poxvirus, encodes a soluble receptor for interferon α/β, which works in a similar manner (Alcami *et al.* 2000). Both human herpesvirus 8 (Kaposi sarcoma-associated virus) and Epstein-Barr virus encode a Bcl-2 homologue which prevents the infected cell from undergoing apoptosis (Henderson *et al.* 1993; Cheng, EH-Y *et al.* 1997). These two gammaherpesviruses also encode proteins (K3 and K5 in KSAV and EBNA-1 in EBV) which downregulate MHC class I expression, thereby decreasing viral antigen presentation on the cell surface. This enables them to prevent cytotoxic T lymphocyte (CTL)-mediated apoptosis of their host cell (Stevenson *et al.* 2000). Herpes simplex virus (HSV) type 1 and type 2 both encode glycoprotein C, which interacts with complement factor C3b, protecting the virus from complement-mediated neutralisation (Kostavasili *et al.* 1997).

1.8.1 Host Immune Modulation by ASFV

Against this background of co-evolution by DNA viruses, is it no surprise that ASFV has also developed a range of proteins to modulate the host immune response, these include:

- **8-DR**, a CD2-homolgue involved in haemaglutination and the attachment of virus to the surface of erythrocytes (Borca *et al.* 1998)
- **A224L**, an inhibitor of apoptosis (IAP)-like protein, which inhibits caspase 3 (Nogal *et al.* 2001; Rodríguez *et al.* 2002)
- **A179L** (5-HL), an antiapoptotic Bcl-2 homologue (Neilan *et al.* 1993; Afonso *et al.* 1996; Brun *et al.* 1996; Brun *et al.* 1998)

- **A238L** (5-EL), an immunomodulatory IκBα/NFAT homologue (Powell *et al.* 1996; Miskin *et al.* 1998; Miskin *et al.* 2000; Tait *et al.* 2000).

The mode of action of 8-DR, A224L and A179L is thought to be relatively straight-forward, however A238L is unusual in this respect. A238L is a novel immunomodulatory protein, 238 amino acids in length, which has been shown *in vitro* to target several key locations within both the porcine NFκB and NFAT immune transcription pathways. It does this by encoding different homologous domains along its length. No other single virus protein has been described which has the versatility to simultaneously target two key immunomodulatory pathways (Miskin *et al.* 2000). A description of these pathways and how A238L interacts with them follows.

1.8.2 NFκB

The transcription factor NFκB (Nuclear Factor kappa B) is highly evolutionarily conserved with homologous proteins existing in plants, insects (*Drosophila*), amphibians (*Xenopus*) and mammals. In mammals, NFκB is found in virtually all cell types where it plays a central role in the regulation of many genes involved in immune and inflammatory responses (Table 1.5). To date, over 300 potential downstream targets have been identified which contain an NFκB-binding motif in their promoter, of which at least half have been demonstrated functionally. Furthermore, NFκB may also act on composite sites in combination with other transcription factors (Kopp and Ghosh 1995; Powell *et al.* 1996; Gilmore 2004).

In addition to immunomodulatory control, NFκB plays an essential role in a range of other cellular processes including proliferation, apoptosis and development. Evidence for these effects are derived from knock-out experiments involving components of the NFκB pathway (Table 1.6) (Ghosh *et al.* 1998; Karin and Ben-Neriah 2000). NFκB is activated in response to a wide variety of signals, including proinflammatory cytokines, lipopolysaccharide (LPS), double-stranded RNA,

Table 1.5

Examples of the wide range of inflammatory and immune system components regulated by NFκB.

Table 1.6

In addition to mediating the immune response, NFκB plays an essential role in a range of cellular processes including proliferation, apoptosis and development; this is evident from the lethality or severe immune and developmental defects seen in knock-out experiments involving components of the NFκB pathway.

Inflammatory component	Examples
Cytokines	IL1, IL2, IL6, IL8, TNFα, TNFβ, IFN-β, IFN-γ
Growth factors	GM-CSF, G-CSF, M-CSF, VEGF
Acute phase proteins	SAA, C-reactive protein, angiotensinogen, tissue factor
Adhesion molecules	ICAM-1, VCAM-1, ECAM-1 (E-selectin)
Immunoreceptors and antigen presentation	MHC class I, MHC class II, TNF receptor
Regulators of apoptosis	Fas, caspase-11, Bcl-2, IAPs
Enzymes	Nitric oxide synthase

Knockout(s)	Phenotype	Reference(s)
p65 (RelA)	Embryonic lethal (E15-16)	(Beg *et al.* 1995b)
p50	General immune defects	(Sha *et al.* 1995)
c-rel	Impaired lymphocyte activation	(Köntgen *et al.* 1995)
RelB	Multiple defects from day 10 post partum	(Burkly *et al.* 1995; Weih *et al.* 1995)
p50, RelB	Lethal < 4 weeks of age	(Weih *et al.* 1997)
p65, p50	Embryonic lethal (E12)	(Horwitz *et al.* 1997)
IκBα	Lethal < 9 days post partum	(Beg *et al.* 1995a; Klement *et al.* 1996)
IKKα	Lethal < 4 hours post partum	(Hu *et al.* 1999; Takeda *et al.* 1999)
IKKβ	Embryonic lethal (E12-14)	(Li, Q *et al.* 1999a; Li, Z-W *et al.* 1999b; Tanaka *et al.* 1999)

ultraviolet light, reactive oxygen intermediates and many other components of physical, metabolic and pathogen-related stress (Gilmore 2004).

NFκB exists as a dimer, the most common form comprising p65 and p50 subunits. All NFκB subunits are members of the rel family of proteins (Figure 1.9). In the quiescent cell, NFκB is located in the cytoplasm bound to its inhibitory protein, IκB (Inhibitor of kappa B) (see 1.8.2.1 below). All rel proteins contain a highly conserved N-terminal 300 amino acid rel homology domain (RHD), responsible for DNA binding, dimerisation, activation and interaction with IκB (Kopp and Ghosh 1995; Foo and Nolan 1999).

1.8.2.1 IκB

The IκB (Inhibitor of kappa B) family is a large group of proteins (Figure 1.9) containing multiple conserved 33 amino acid ankyrin repeats. These stacked helical domains are involved in protein-protein interaction and mediate binding to the RHD of NFκB subunits. The IκB family consists of IκBα, IκBβ, IκBγ, IκBε and Bcl3. Different members of the IκB family have affinity for specific NFκB dimers. All IκB family members inhibit NFκB-dependent transcription except Bcl-3 which enhances NFκB activity by binding p50 and p52 homodimers. These homodimers are transcriptionally inactive and would otherwise compete with active forms of NFκB (Foo and Nolan 1999). IκBα is the most abundant IκB protein and inhibits the common p65-p50 NFκB heterodimer (Kopp and Ghosh 1995). When the NFκB pathway is activated, IκBα is rapidly phosphorylated at two serine residues (Ser32 and Ser36) by the IκB kinase complex (IKK), ubiquinated (at residues Lys21 and Lys22) and targeted for degradation via the 26S proteasome (Roff *et al.* 1996; Foo and Nolan 1999; Karin 1999; Zandi and Karin 1999; Karin and Ben-Neriah 2000). This exposes the nuclear localisation signal of NFκB and allows it to translocate to the nucleus where it upregulates transcription by binding to a 10bp consensus κB binding site (5'-GGGRNNYYCC-3') (Figure 1.10a). IκBα is among the many

Figure 1.9 – The Rel, IκB and NFAT Protein Families

The members of the Rel, IκB and NFAT families of proteins are illustrated (adapted from diagrams by Ghosh *et al.* (1998), Hogan *et al.* (2003) and Macián *et al.* (2000)). Members of these families contain a combination of highly conserved domains, including the rel homology domain (RHD), rel homology region (RHR), regulatory domain and ankyrin repeats. Some family members also contain a transactivation domain (TD), leucine zipper (LZ), glycine-rich region (GRR) and/or a serine-rich region (SRR). Arrows relate to internal proteolytic cleavage points in p100 and p105, which when cleaved become p52 and p50 respectively.

Dif, Dorsal and Relish are from *Drosophila* and RelB, c-rel, p65, p100/p52 and p105/p50 are mammalian rel proteins. In the IκB family, IκBγ, Bcl-3, IκBε, IκBα and IκBβ are found in higher vertebrates, whereas Cactus is from *Drosophila*. ASFV-encoded A238L has been included here as an IκB-homologue. NFAT proteins 1-4 are the classical family members, containing both a regulatory domain and the RHR, whereas as NFAT 5 is a primordial family member, containing only the RHR, and is the only NFAT protein to also be represented in the *Drosophila* genome. It bears the most resemblance to the rel proteins and does not co-operatively bind with AP-1.

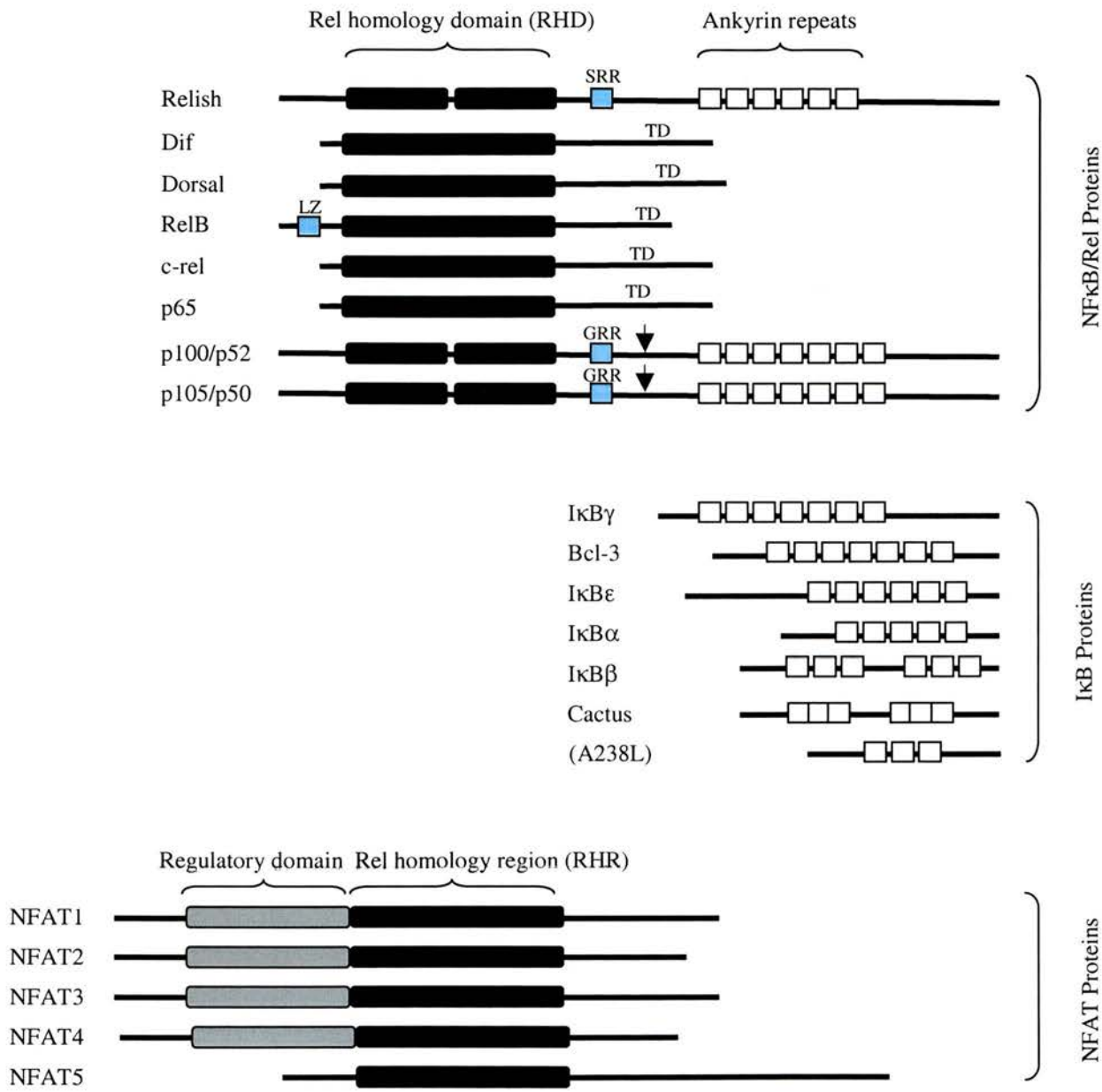

Rel homology domain (RHD)
Ankyrin repeats
SRR
Relish
Dif
TD
Dorsal
TD
LZ
RelB
TD
c-rel
TD
p65
TD
GRR
p100/p52
GRR
p105/p50
NFκB/Rel Proteins
IκBγ
Bcl-3
IκBε
IκBα
IκBβ
Cactus
(A238L)
IκB Proteins
Regulatory domain
Rel homology region (RHR)
NFAT1
NFAT2
NFAT3
NFAT4
NFAT5
NFAT Proteins

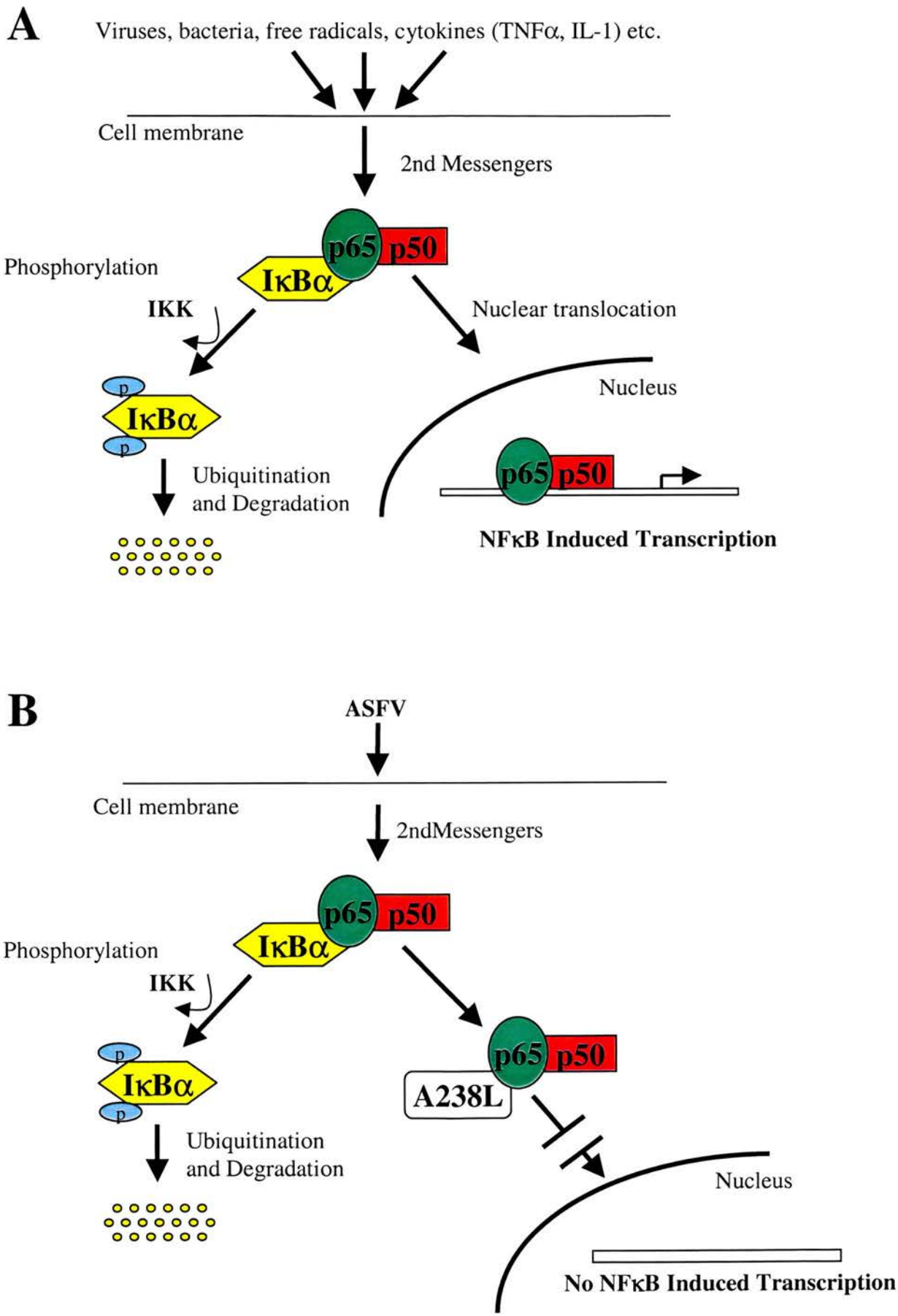
Figure 1.10 – NFκB and its proposed interaction with A238L during ASFV infection

A NFκB in the Normal Cell

In the quiescent cell, NFκB (p65-p50 heterodimer) is located in the cytoplasm, bound to its inhibitor, IκBα. When the cell is activated, IκBα is rapidly phosphorylated by the IκB kinase complex (IKK), ubiquinated and targeted for degradation via the 26S proteasome. The active NFκB complex translocates to the nucleus where it upregulates transcription of genes involved in the immune and inflammatory response (Kopp and Ghosh 1995; Foo and Nolan 1999; Karin 1999).

B NFκB in the ASFV-infected Cell

When activated during ASFV infection, the NFκB complex dissociates from IκBα in the usual way. However, it has been proposed that the ASFV-encoded A238L binds the p65 subunit of NFκB, behaving as an IκBα homologue. This prevents nuclear translocation of NFκB and upregulation of the immune response (Powell *et al.* 1996; Revilla *et al.* 1998; Tait *et al.* 2000).


A
Viruses, bacteria, free radicals, cytokines (TNFα, IL-1) etc.
Cell membrane
2nd Messengers
p65
p50
IκBα
Phosphorylation
IKK
Nuclear translocation
p
IκBα
p
Nucleus
Ubiquitination
and Degradation
p65
p50
NFκB Induced Transcription
B
ASFV
Cell membrane
2ndMessengers
p65
p50
IκBα
Phosphorylation
IKK
p
IκBα
p
p65
p50
A238L
Ubiquitination
and Degradation
Nucleus
No NFκB Induced Transcription

downstream targets of NFκB. This acts to replenish degraded IκBα and inhibit NFκB in an autoregulatory manner (de Martin *et al.* 1993; Cheng, Q *et al.* 1994).

1.8.2.2 IκBα-independent Regulation by Phosphorylation

In addition to primary IκB-mediated control of NFκB, a secondary level of control has been more recently identified. Although all NFκB subunits are constitutively phosphorylated, the p65 subunit can also be further phosphorylated in a signal-induced, signal-specific manner. NFκB activation due to 'typical' agents, such as proinflammatory cytokines, results in phosphorylation of p65 at serine residues 276, 529 and 536, which significantly increases the transcriptional activity of NFκB (Wang and Baldwin Jr. 1998; Schmitz *et al.* 2001; Vermeulen *et al.* 2002). Phosphorylation of Ser276 (located in the rel homology domain) by protein kinase A (PKA), has been shown to allow binding of the co-activator CREB-binding protein (CBP)/p300 (Zhong *et al.* 1998; Zhong *et al.* 2002). This interaction is believed to help link NFκB with the general transcription machinery, increasing its binding potential (Schmitz *et al.* 2001). Indeed, overexpression of CBP has been shown to enhance NFκB transcription (Gerritsen *et al.* 1997; Perkins *et al.* 1997).

Ser529 and Ser 536 lie within the transactivation domain 1 (TA1), which comprises the C-terminal 30 amino acids of p65 (Schmitz *et al.* 1995). Ser529 phosphorylation is mediated by casein kinase II (CKII) (Bird *et al.* 1997; Wang *et al.* 2000), whereas Ser536 is phosphorylated by the IKK complex (Sakurai *et al.* 1999). Phosphorylation of Ser529 and Ser536 does not appear to alter DNA binding or nuclear translocation but increases the transcriptional activity of NFκB by an unknown mechanism (Wang and Baldwin Jr. 1998; Wang *et al.* 2000).

When NFκB is activated by 'atypical' agents, for example UV-C and some chemotherapeutic compounds (e.g. daunorubicin), there is little NFκB transcriptional activity despite a strong induction of DNA binding. Furthermore, this association actively represses any subsequent activation with 'typical' agents. This may be due

to an increase in association of p65 with the class I histone deacetylase complexes (HDACs 1, 2 and 3), which act to repress transcriptional activity by deacetylation of the histone proteins round which the target promoter is wrapped (Campbell *et al.* 2004) (see Information Box 1.2). This mechanism may explain why NFκB activation can on some occasions mediate an anti-proliferative and/or proapoptotic response.

1.8.2.3 The ASFV-encoded Protein, A238L, Interacts with the NFκB Pathway

A238L shares 40% homology and 21% identity with domestic pig IκBα and binds NFκB (p65-p50) in place of IκBα (Tait *et al.* 2000). It has regions which are homologous to three out of the five conserved ankyrin repeats in IκBα, involved in binding to NFκB (de Martin *et al.* 1993; Powell *et al.* 1996). A238L specifically co-precipitates with the p65 subunit and selectively inhibits proinflammatory gene upregulation by confining NFκB to the cytoplasm (Figure 1.10b). In addition, anti-inflammatory gene expression (TGFβ) has been shown to be upregulated during ASFV infection (Powell *et al.* 1996). A238L is believed to effectively act as an 'NFκB super-repressor', as it lacks the two serine residues homologous to Ser32 and Ser36 and is therefore not susceptible to signal-induced phosphorylation, ubiquitination and degradation (Powell *et al.* 1996; Revilla *et al.* 1998; Tait *et al.* 2000). ASFV infection initially stimulates the NFκB pathway, promoting degradation of IκBα, which enables A238L to bind NFκB in its place (Tait *et al.* 2000). No 'tri-molecular' complexes of p65, A238L and IκBα have been detected, suggesting that A238L and IκBα bind NFκB at the same or overlapping sites (Tait *et al.* 2000).

A238L can be detected in both 28kDa and 32kDa forms in the cell. Only the 32kDa form appears to interact with p65. This variation in molecular weight may be due to post-translational modification (Tait *et al.* 2000).

Information Box 2: Histones and HDACs

DNA is packaged into highly organised structures called chromatin. Chromatin is made of subunits termed nucleosomes, which each comprise 146bp DNA wrapped around an octamer of histone proteins (histones). Histones are highly conserved, but can vary markedly in their posttranslational modification. The N-terminal tail of each histone protrudes from the nucleosome and may be modified by acetylation, phosphorylation, methylation, sumoylation and ubiquitination. The nature of these modifications can have a profound effect on the structure and pattern of chromatin condensation and gene expression. This is known as the histone code (Jenuwein and Allis 2001; Turner 2002; Iizuka and Smith 2003).

Acetylation of histones occurs on lysine residues and is brought about by histone acetyltransferases (HATs); this enables an increase in transcriptional activity. Conversely, histone deacetylase complexes (HDACs) reduce levels of acetylation and cause a decrease in gene expression (Berger 2002; Thiagalingam *et al.* 2003).

1.8.3 NFAT

The second transcription pathway with which A238L has been shown to interact is NFAT (Nuclear Factor of Activated T cells). NFAT proteins are expressed in many classes of immune cell. They are a family of transcription factors central to upregulating the expression of components involved in the immune and inflammatory response (Rao *et al.* 1997). Sequence analysis has identified around 50 genes containing possible NFAT response elements. Twelve have been confirmed experimentally (Kel *et al.* 1999). The distinguishing feature of NFAT is its regulation by Ca^{2+} and the Ca^{2+}/calmodulin-dependent serine phosphatase, calcineurin (see 1.8.3.1 below). Stimulation of immunoreceptors and/or receptor tyrosine kinases (RTK) on the cell surface results in the release of intracellular Ca^{2+} from the endoplasmic reticulum. This causes calmodulin to bind calcineurin, which in turn activates NFAT; thus providing a direct link between Ca^{2+} signalling and gene expression (Hogan *et al.* 2003).

There are five members of the NFAT family (Figure 1.9). The classical NFAT proteins 1-4 share two regions of particular homology: The regulatory domain and the rel homology region (RHR). In contrast, NFAT5 is the primordial family member and only contains the RHR. Its mode of action is unusual and poorly understood. Therefore this introduction will consider only NFAT proteins 1-4. The RHR contains two motifs clearly homologous to the rel homology domain (RHD) of rel proteins. Hence, NFAT is considered an additional member of an extended rel protein family (Macián *et al.* 2000; Hogan *et al.* 2003). The regulatory domain is serine-rich, highly phosphorylated and binds calcineurin using two distinct docking domains (Figure 1.11a).

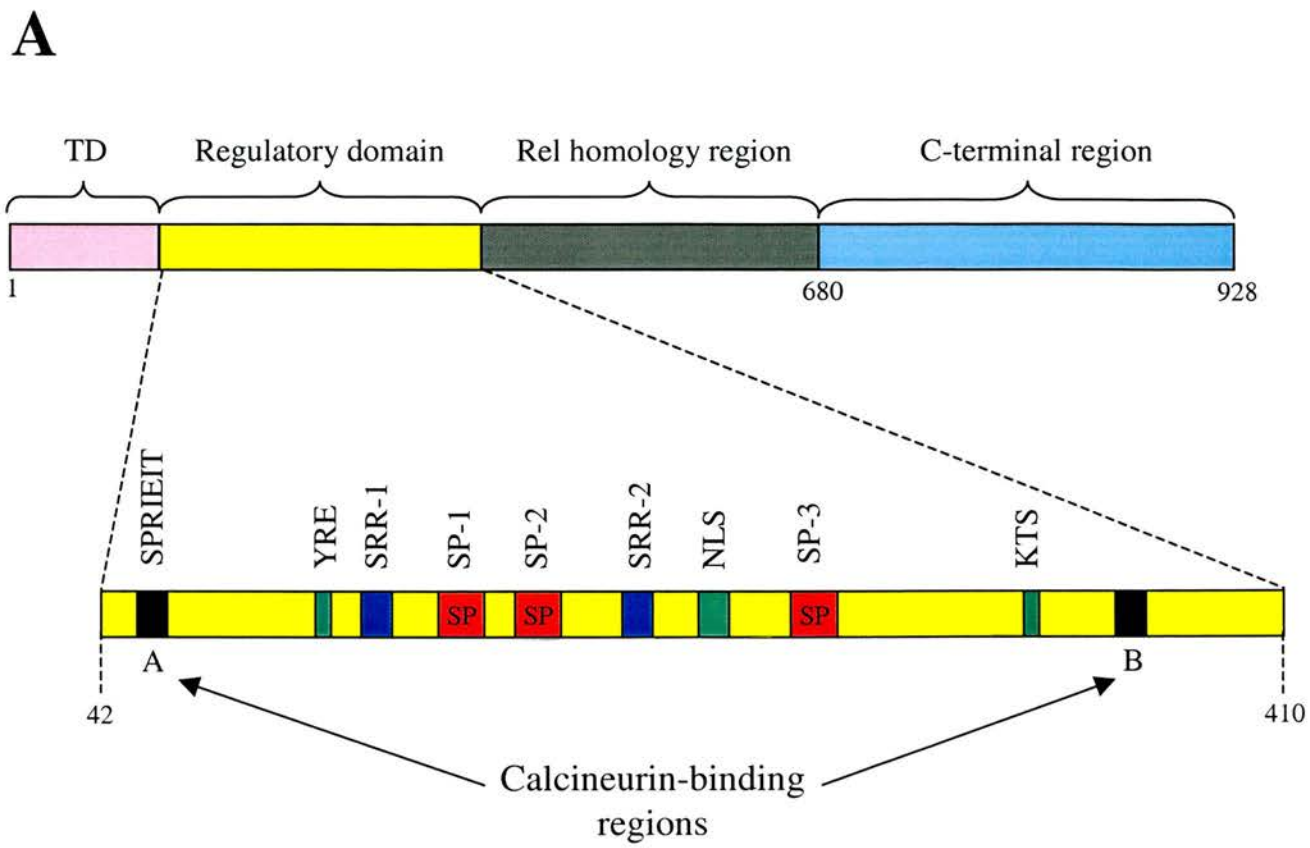
1.8.3.1 Calcineurin A

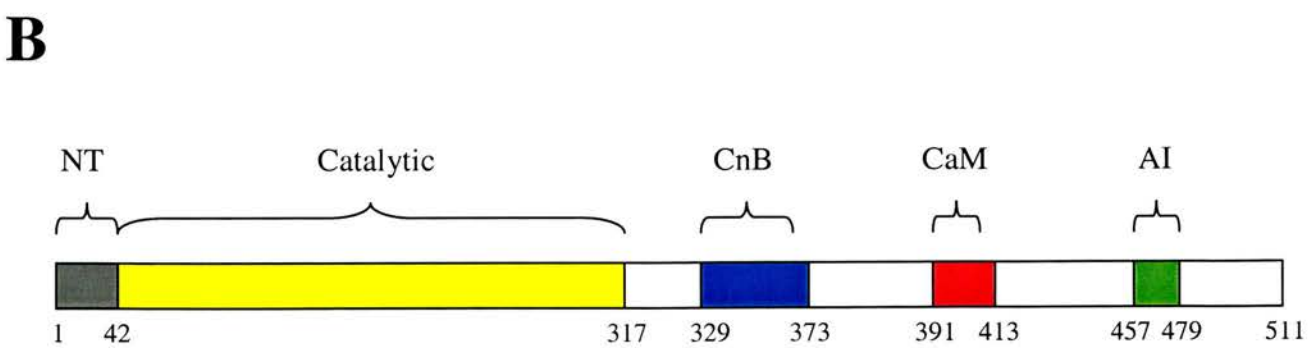
Calcineurin (protein phosphatase 2B) is a member of the serine/threonine protein phosphatase family. Calcineurin is a heterodimer, consisting of calcineurin A and

Figure 1.11 – NFAT Regulatory Domain and Calcineurin A

A NFAT and its regulatory domain (adapted from illustrations by Macián et al. (2000) and Hogan et al. (2003)). NFAT proteins can be divided into four regions, an N-terminal transactivation domain (TD), the regulatory domain, the rel homology region and the C-terminal region. The regulatory domain contains a number of highly conserved sequence motifs. These include two regions responsible for binding calcineurin, a YRE motif, two serine-rich regions (SRR), three serine-proline repeat regions (SP), the nuclear localisation signal KRR (NLS) and a KTS motif. Sequence is numbered by amino acid in murine NFAT1.

B Diagrammatic representation of Calcineurin A (adapted from Rusnak and Mertz (2000)). Calcineurin A encodes five functional domains: an N-terminal domain (NT), the catalytic domain, the calcineurin B-binding domain (CnB), the calmodulin-binding domain (CaM) and an autoinhibitory domain (AI). Sequence is numbered by amino acid in rat Calcineurin Aα.


A
TD
Regulatory domain
Rel homology region
C-terminal region
1
680
928
SPRIEIT
YRE
SRR-1
SP-1
SP-2
SRR-2
NLS
SP-3
KTS
SP
SP
SP
A
B
42
410
Calcineurin-binding
regions


B
NT
Catalytic
CnB
CaM
AI
1
42
317
329
373
391
413
457
479
511

calcineurin B subunits (Rao *et al.* 1997; Rusnak and Mertz 2000). Calcineurin A has a catalytic domain at the N-terminus (50% of the gene) and three regulatory domains at the C-terminus: the calcineurin B-binding domain, the calmodulin-binding domain and the autoinhibitory domain (Rusnak and Mertz 2000) (Figure 1.11b). In the event of a rise in intracellular Ca^{2+} concentration (hypertonic stress), calmodulin binds calcineurin A causing a conformational change, which shifts the autoinhibitory domain away from the catalytic site. Serine-rich motifs in the regulatory domain of NFAT are then dephosphorylated by calcineurin, which unmasks the nuclear localisation sequence and allows translocation (Figure 1.12a). DNA binding occurs as either NFAT homodimers at κB-like motifs or cooperatively with the unrelated AP-1 (Fos-Jun) transcription factor at composite NFAT/AP-1 elements (Figure 1.13). As AP-1 is modulated by a different array of signals, including protein kinase C (PKC) and mitogen-activated protein (MAP) kinases, NFAT/AP-1-mediated transcription represents an integration of a wide variety of signalling inputs (Rao *et al.* 1997; Chen *et al.* 1998; Macian *et al.* 2000; Macián *et al.* 2000; Rusnak and Mertz 2000; Hogan *et al.* 2003).

In man, there are three isoforms of calcineurin A encoded by different genes on different chromosomes: Calcineurin Aα (*PPP3CA* gene, chromosome 4), calcineurin Aβ (*PPP3CB* gene, chromosome 10) and calcineurin Aγ (*PPP3CC* gene, chromosome 8) (Ensembl Genome Browser; http://www.ensembl.org/). In the rat, calcineurin Aβ (CnAβ) is most commonly expressed in the spleen, thymus and lymphocytes (Rusnak and Mertz 2000). Indeed, CnAβ is the isoform predominantly expressed in porcine macrophages during ASFV infection and was found to bind NFAT2 during yeast-two hybrid studies (Miskin *et al.* 2000).

1.8.3.2 The ASFV-encoded Protein, A238L, Interacts with the NFAT Pathway

The ASFV-encoded protein, A238L, interacts with both calcineurin Aβ and cyclophilin A. This was demonstrated by yeast two-hybrid experiments using A238L

Figure 1.12 – NFAT and its Proposed Interaction with A238L during ASFV Infection

A NFAT in the Normal Cell

In the quiescent cell, NFAT is inactive, phosphorylated and located in the cytoplasm. However, a rise in intracellular Ca^{2+} concentration stimulates the binding of calmodulin (CaM) to calcineurin. This causes a change in conformation of the calcineurin A (CnA) subunit which then dephosphorylates NFAT. This active form of NFAT translocates to the nucleus and upregulates expression of genes involved in the immune response (Rao *et al.* 1997; Rusnak and Mertz 2000).

B NFAT in the ASFV-infected Cell

During ASFV infection, it has been proposed that A238L prevents activation of NFAT by two mechanisms (Miskin *et al.* 1998; Miskin *et al.* 2000):

1. A238L binds to the immunophilin carrier cyclophilin A (CypA) and inhibits calcineurin Aβ function, thereby preventing dephosphorylation and activation of NFAT.

2. A238L binds calcineurin Aβ directly, using a calcineurin-docking domain homologous to that found on NFAT proteins. In this manner, A238L competes with calcineurin for binding to NFAT and prevents its dephosphorylation and activation.

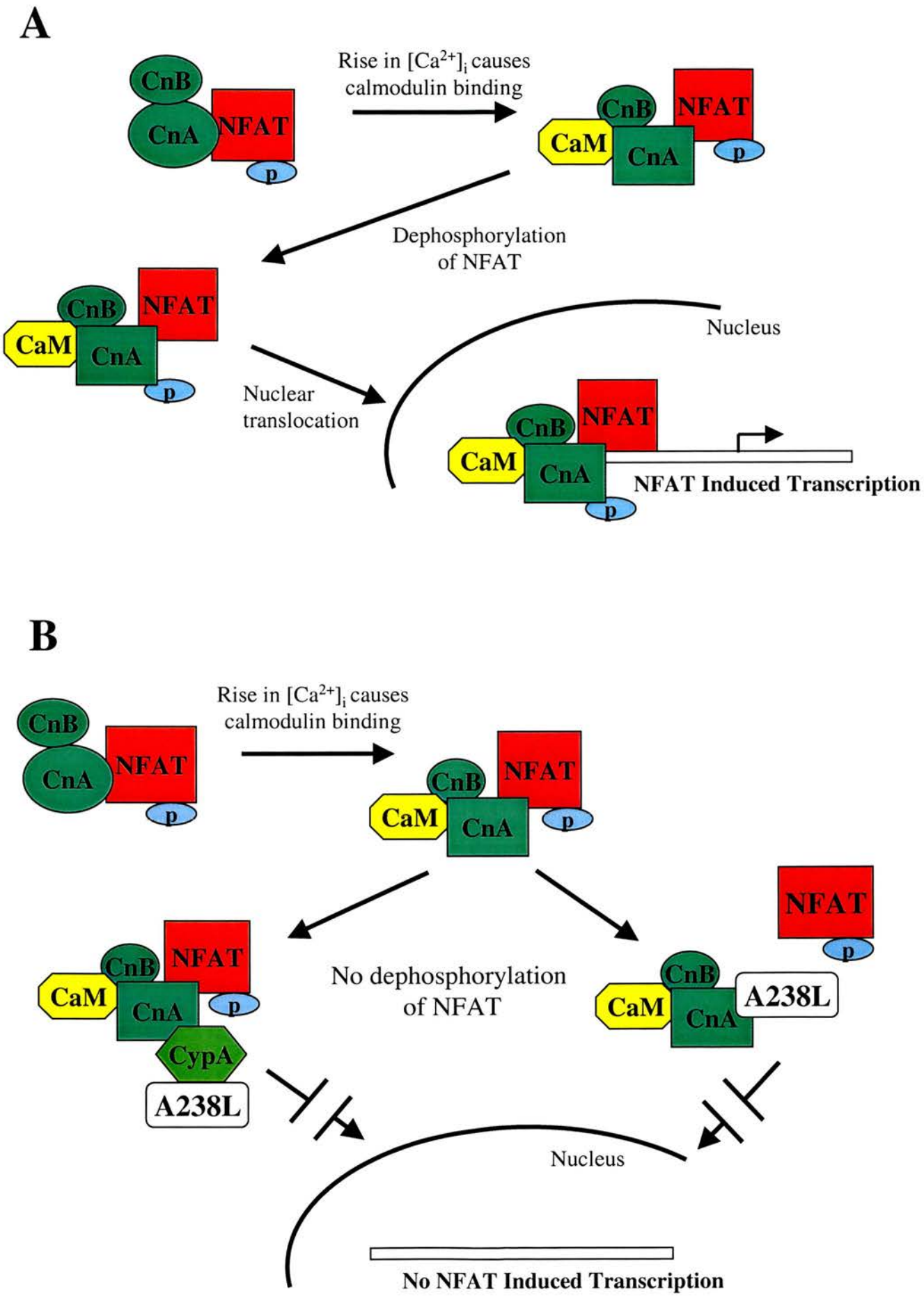
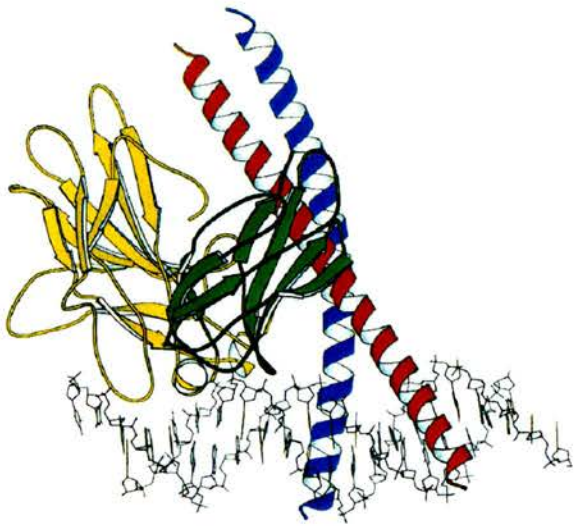

A
CnB
CnA
NFAT
p
Rise in [Ca2+]i causes calmodulin binding
CaM
Dephosphorylation of NFAT
Nucleus
Nuclear translocation
NFAT Induced Transcription
B
Rise in [Ca2+]i causes calmodulin binding
No dephosphorylation of NFAT
CypA
A238L
Nucleus
No NFAT Induced Transcription

Figure 1.13 – Co-operative Binding of NFAT and AP1 to DNA

All NFAT proteins (with the exception of NFAT5) can bind DNA either as a homodimer or cooperatively with the transcription factor activator protein 1 (AP1). AP1 consists of Fos and Jun subunits (Chen, L *et al.* 1998b; Hogan *et al.* 2003). This diagram is taken from Chen *et al.* (1998a) and illustrates NFAT1 (rel homology region yellow/green) cooperatively with AP1 (Fos, red; Jun, blue) to the antigen receptor response element 2 (ARRE2) sequence of the interleukin 2 promoter.



5'-TTGGAAAATTTGTTTCATAG -3'
3'- CCTTTTAAACAAAGTATCAA-5'

NFAT1/Fos-Jun/ARRE2

as 'bait' to screen a cDNA library from porcine alveolar macrophage. Therefore, A238L is proposed to inhibit activation of NFAT by two mechanisms (Figure 1.12b):

1. By binding to the immunophilin carrier, cyclophilin A (in a similar manner to the immunosuppressive drugs cyclosporin A and FK506), to form a complex which in turn binds calcineurin A and prevents it from dephosphorylating NFAT (Schreiber and Crabtree 1992; Kissinger *et al.* 1995; Miskin *et al.* 1998).

2. By binding directly to calcineurin Aβ using an eight amino acid calcineurin-binding domain (PxIxITxC/S), homologous to that present on NFAT proteins. This prevents dephosphorylation and subsequent activation of NFAT (Miskin *et al.* 1998).

The binding site for A238L on calcineurin Aβ has only been loosely localised to between residues 32-350. This region includes part of the catalytic domain and the calcineurin B binding domain. Interestingly, it is the 28kDa form of A238L which has been shown to bind calcineurin Aβ and cyclophilin A in the NFAT pathway (Miskin *et al.* 1998; Miskin *et al.* 2000; Tait *et al.* 2000), compared to the 32kDa form interacting with NFκB subunit p65.

1.8.4 Summary: Role of A238L in Disease Pathogenesis

It is clear that the ASFV-encoded protein A238L represents a novel and potent immunoregulatory mechanism by virtue of its ability to modulate both NFκB- and NFAT-mediated transcription. As a result, the host proteins with which A238L is known to interact, calcineurin Aβ, cyclophilin A and p65, and the proteins it mimics, IκBα and NFAT2, represent key targets for investigation.

The variation in the qualitative and quantitative features of the pathology seen during ASFV infection of different pig species may reflect a variable ability of the virus to effectively modulate the host immune response. Alternatively, it may indicate a

fundamental evolutionary distinction between the immune responses of the warthog compared to the domestic pig.

1.9 Hypothesis and Objectives

It is hypothesised that variation between the warthog and domestic pig TNFα promoter, the cargo-specific tether light chain dynein and the A238L-targeted/mimicked proteins, play a role in determining the different pathophysiological outcomes of ASFV infection.

To test this hypothesis, comparative sequence analysis of ASFV-susceptible domestic pig, resistant warthog and phenotypically unknown babirusa light chain dynein, cyclophilin A, NFAT2, IκBα, calcineurin Aβ and p65 nucleotide and primary protein sequence has been undertaken. The extent and degree of sequence variation between these species has been used to identify gene regions capable of modifying interaction with A238L and/or modulating the immune response and in doing so, suggest putative regions which may confer host resistance or susceptibility to ASF. In addition, the TNFα promoter has been sequenced in these species and assessed for polymorphisms which may impact on its activity.

Functionality attached to sequence differences identified between the warthog and domestic pig TNFα promoter and p65 has been assessed *in vitro*, by use of luciferase reporter constructs.

2 Materials and Methods

This chapter provides a description of routine protocols. Additional information is supplied where methods require further explanation or have been adapted from standard protocols provided by the manufacturer. Precise cloning strategies and experimental approaches developed to address specific questions are given in the relevant chapters.

2.1 Sample Collection

Samples from a domestic pig were collected on 19th February 2002 at the Institute for Animal Health (Pirbright), samples from a female warthog were collected on 28th March 1999 at Rotterdam Zoo (Holland) and samples from a female babirusa were collected on 20th May 1999 at Marwell Zoo (Winchester). Samples were stored and processed as described in the relevant sections below.

2.2 Laboratory Resources

Details of reagents, chemicals, kits and suppliers are given in Table 2.1 and recipes/solutions in Table 2.2.

2.3 Molecular Biological Techniques: General

2.3.1 Nucleic Acid Extraction

2.3.1.1 DNA Isolation

Genomic DNA isolations were carried out using components from the Nucleon BACC2 Extraction Kit for Blood and Cell Cultures (Nucleon Biosciences).

Table 2.1

Reagents, chemicals, kits and suppliers

Item	Supplier
Agar, select	Sigma
Agarose MP	Roche Diagnostics
Alkaline phosphatase, shrimp	Roche Diagnostics
Ampicillin (100mg/ml)	Sigma
Aquasan Tablets	Inverclyde Biochemicals
BCA Protein Assay Kit	Pierce
Bromophenol blue (0.25%)	Sigma
Carbon dioxide (gas)	BOC
Deoxynucleotide Set	Roche Diagnostics
DMSO	Sigma
DNA Molecular Weight Markers (II, XIV and XVII)	Roche Diagnostics
Dolbecco's Modified Eagle Medium (DMEM)	Sigma
Dry ice (solid carbon dioxide)	
Dual-Reporter Luciferase Assay System	Promega
EDTA	Sigma
Ethanol	Sigma
Ethidium bromide	Sigma
Expand *High Fidelity* PCR System	Roche Diagnostics
Ficoll (type 400) (15%)	Sigma
Foetal Calf Serum (FCS)	Invitrogen/Gibco
G418 (50mg/ml)	Sigma
Glacial acetic acid	Sigma
Glucose	Sigma
IPTG	Sigma
Haz-Tabs	
L-Glutamine (LG)	Sigma
Lipofectamine 2000 Reagent	Invitrogen
Magnesium chloride	Sigma
mRNA Isolation Kit for Blood/Bone Marrow	Roche Diagnostics
Nitrogen (liquid)	
Nucleon BACC2 DNA Extraction Kit	Nucleon Biosciences
Oligonucleotides	Qiagen (Operon)
pBudCE4.1 Vector	Invitrogen
pcDNA3	Invitrogen
pcDNA3.1/Zeo+	Invitrogen
PCR-grade water	Sigma
pFLAG-CMV-4 Vector	Sigma
pGEM™-T Easy Vector System	Promega
pGL3-Basic Vector	Promega
phRL-null Vector	Promega
Phosphate-buffered saline (10x)	Invitrogen
pRL-TK Vector	Promega
Propan-2-ol (isopropanol)	Sigma
Proteinase K	Roche Diagnostics
Protein G Sepharose	Amersham
Qiagen Plasmid Maxi Kit	Qiagen

QIAprep™ Spin Miniprep Kit	Qiagen
QIAquick™ Gel Extraction Kit	Qiagen
QIAquick™ PCR Purification Kit	Qiagen
QIAshredder	Qiagen
Rapid DNA Ligation Kit	Roche Diagnostics
Restriction Endonucleases	Roche Diagnostics
RNA/DNA Stabilization Reagent for Blood/Bone Marrow	Roche Diagnostics
Select Yeast Extract	Sigma
Sensiscript RT Kit	Qiagen
SL-8502 BACC2 (DNA Extraction Kit)	Nucleon Biosciences
SMART™ RACE cDNA Amplification Kit	BD Biosciences Clontech
Sodium chloride	Sigma
Tris	Sigma
Trypsin-EDTA (TE)	Sigma
Tryptone	Sigma
X-GAL Tablets	Roche Diagnostics
XL-1 Blue Subcloning-Grade Competent Cells	Stratagene
Xylene cyanol (0.25%)	Sigma
Zeocin (100mg/ml)	Cayla

Table 2.2

Solutions and recipes

DNA Loading Buffer	0.025g Bromophenol Blue (0.25%) 0.025g Xylene cyanol (0.25%) 1.5g Ficoll (type 400) (15%) Make up to 10ml with distilled water
L-Broth	10g tryptone 5g yeast extract 10g Sodium chloride Make up to 1 litre with distilled water
L-Broth Agar	500ml L-Broth (see above) 7.5g Agar
SOB Medium	20g tryptone 5g yeast extract 0.5g Sodium chloride Make up to 1 litre with distilled water
SOC Medium	10ml SOB (see above) 200µl Glucose (1M) 50µl Magnesium chloride (2M)
Homogenisation Buffer	140mM NaCl 50mM Tris HCl (pH 7.5) 1mM EDTA 1% Triton Protease inhibitors: (aprotinin, 2ug/ml; antipain 1ug/ml; leupeptin, 0.2mg/ml; phenylmethanesulphonyl fluoride (PMSF), 1mM).
5x TAE	121g Tris(hydroxymethyl)aminomethane (Tris) 28.55ml Glacial acetic Acid 50ml EDTA (0.5M) Make up to 5 litres with distilled water
Tris-buffered saline (TBS)	100mM Tris HCl (pH7.5) 150mM NaCl (0.9%)
TBST	TBS (see above) 0.1% Tween 20
Western Blot Running Buffer	Outer chamber: 1x NuPAGE MOPS SDS Running Buffer (Invitrogen) Inner chamber: 1x NuPAGE MOPS SDS Running Buffer (Invitrogen) 0.25% NuPAGE Antioxidant (Invitrogen)
Western Blot Transfer Buffer	Inner chamber only: 1x NuPAGE Transfer Buffer (Invitrogen) 0.001% NuPAGE Antioxidant (Invitrogen) 20% Methanol

DNA was isolated from skeletal muscle tissue stored at -70°C in 20% DMSO/saturated salt (NaCl) solution. For each extraction, a 0.2g piece of tissue was removed using a sterile scalpel blade and rinsed in 5ml sterile water to remove excess salt solution. The sample was then wrapped in two layers of aluminium foil, submerged in liquid nitrogen and pulverised with a hammer. The pulverised tissue was scraped into a 1.5ml microfuge tube containing 450µl Reagent B (cell lysis buffer) and further ground using a plastic homogeniser. Proteinase K (20µl of 20mg/ml) was added to the sample and incubated overnight in a waterbath at 50°C. The following morning the sample was spiked with a further 100µg Proteinase K and returned to the waterbath for one hour. The sample was then heated to 65°C for one hour to denature the Proteinase K and centrifuged at full speed for five minutes to pellet the cell debris. The supernatant was removed to a clean tube and deproteinised using 128µl 5M sodium perchlorate solution; then 750µl chloroform was added and the tube placed on a roller-mixer for 30 minutes at room temperature to emulsify the phases. 100µl Nucleon resin was gently added to each tube and centrifuged at 9520 x g (10 000rpm) for 20 minutes. The upper phase containing DNA (situated above the resin layer) was carefully removed into a clean tube using a Pasteur pipette. To this, 900µl cold 100% ethanol (2 volumes) was added and the tube inverted several times to precipitate the DNA. The tube was then centrifuged at full speed for five minutes to pellet the DNA and the supernatant removed and replaced with 900µl cold 70% ethanol. Once again the tube was inverted several times and centrifuged at full speed for five minutes. The 70% ethanol was removed and the DNA pellet allowed to air-dry for ten minutes before redissolving in 350µl 1x Tris-EDTA (TE) (pH 8.4).

2.3.1.2 mRNA Isolation

RNAse-free practice was strictly observed at all times, including the use of certified RNAse-free plasticware and microfuge tubes. All glassware and metal spatulas were wrapped in foil and baked at 240°C for 6 hours to degrade any RNAses. Folded pieces of aluminium foil were also baked, which when opened out served as an RNAse-free surface to work on. One set of pipettes was kept solely for RNA work and these were swabbed with RNAzap (Qiagen) before use. An RNA-dedicated

section of the bench was cleaned with 0.1% diethylpyrocarbonate (DEPC)-treated water and wiped down with RNAzap. Latex gloves were treated with RNAzap at regular intervals.

Whole blood (5ml) was collected into EDTA and gently mixed. This anticoagulated blood was immediately added to 50ml DNA/RNA Stabilisation Reagent for Blood/Bone Marrow (Roche Diagnostics) and stored at –70°C. The stabilised blood was processed using the associated mRNA Isolation Kit for White Blood Cells (Roche Diagnostics) according to the manufacturer's standard protocol.

RNA concentration was determined using a UV spectrophotometer. In addition, 1µl was run on a 1% agarose gel to check its integrity. The 28S and 16S ribosomal bands were bright and sharp with limited smearing between them (the presence of smearing denotes degradation).

2.3.2 cDNA Synthesis

Approximately 1µg mRNA was used as a template for cDNA synthesis in a total reaction volume of 20µl reverse transcription mix comprising 1xBuffer RT containing a 2mM mix of the four deoxynucleoside triphosphate bases (dNTPs) (0.5mmol dATP, dTTP, dCTP and dGTP)); 1µM oligo-dT_{25} primer; and 1µl *Sensiscript*™ Reverse Transcriptase (Qiagen). The samples were incubated in a waterbath at 37°C for one hour.

2.3.3 Polymerase Chain Reaction (PCR) - Standard

Polymerase Chain Reaction (PCR) was performed using 1µl of cDNA as template in 25µl reaction mixture: 1xPCR buffer containing a 2mM dNTPs, 2.0mM $MgCl_2$, 20pmol of each PCR oligonucleotide primer and 0.7U *High Fidelity* DNA polymerase (Roche Diagnostics). The PCR programme comprised an initial denaturation step of 3 minutes at 94°C followed by 35 cycles, each consisting of a

denaturation, annealing and extension stage. This was concluded by a final extension stage for 10 minutes at 72°C. The temperature and duration of each stage of the 35 cycles was optimised for each individual target sequence.

Each 25µl PCR product was mixed with 3µl loading buffer and run on a 1% agarose gel containing 2µg/µl ethidium bromide (EtBr). Product size was verified by comparison with DNA molecular weight marker II, (λ *Hin*d III digest), XIV (100bp ladder) or XVII (500bp ladder) as appropriate (Roche Diagnostics).

2.3.4 Hot Start/Touchdown PCR

For some complex PCR templates (e.g. DNA or GC-rich sequence) a combination of a hot-start and a touchdown cycle was utilised. This process denatures the template, prevents non-specific binding during the early cycles, and activates the polymerase. For these reactions, 50ng of genomic DNA was used as template in a total reaction volume of 25µl: 1xPCR buffer (*GeneAmp*, Perkin Elmer) containing 2mM dNTPs, 2.5mM $MgCl_2$, 0.5U *AmpliTaq* Gold DNA polymerase (Perkin Elmer) and 20pmol of each oligonucleotide primer. A 36-cycle touchdown PCR programme was used, comprising an initial 95°C denaturation for twelve minutes (the hot start required to activate the polymerase and completely denature the template) followed by: two cycles with a 63°C annealing phase, two cycles with a 62°C annealing phase and two cycles with a 61°C annealing phase. This was followed by 30 cycles with a 59°C annealing phase and a final 72°C extension for ten minutes (See Table 3.2, TNFα promoter as an example). This PCR product was run out on a 1% agarose gel as previously described.

2.3.5 SMART RACE PCR

To amplify unknown sequence which occurred 5' and 3' to a region of known sequence, a technique termed Rapid Amplification of cDNA Ends (RACE) was employed. For this purpose, the SMART™ (Switching Mechanism At 5' end of

RNA Transcript) RACE cDNA Amplification Kit (BD Biosciences Clontech) was used to create 5' and 3' RACE products from extracted RNA. These provide full-length cDNA transcripts of every mRNA in the sample, which contain a short adapter of known sequence at either the 5' or 3' end.

5' and 3' 'RACE-ready' cDNA libraries were produced using *Powerscript*™ reverse transcriptase and the SMART IIA™ oligonucleotide (BD Biosciences Clontech). Following reverse transcription of polyadenylated mRNA using a standard oligo-dT_{25} primer, the SMART IIA adapter was ligated to the 5' end of 5' RACE-ready cDNA (Figure 2.1a). To create 3' RACE-ready cDNA, the oligo-dT primer also contained the SMART adapter sequence, which became incorporated into the 3' end of the cDNA transcript (Figure 2.1b). These cDNA libraries were diluted in a Tricine-EDTA buffer and used in subsequent PCR reactions.

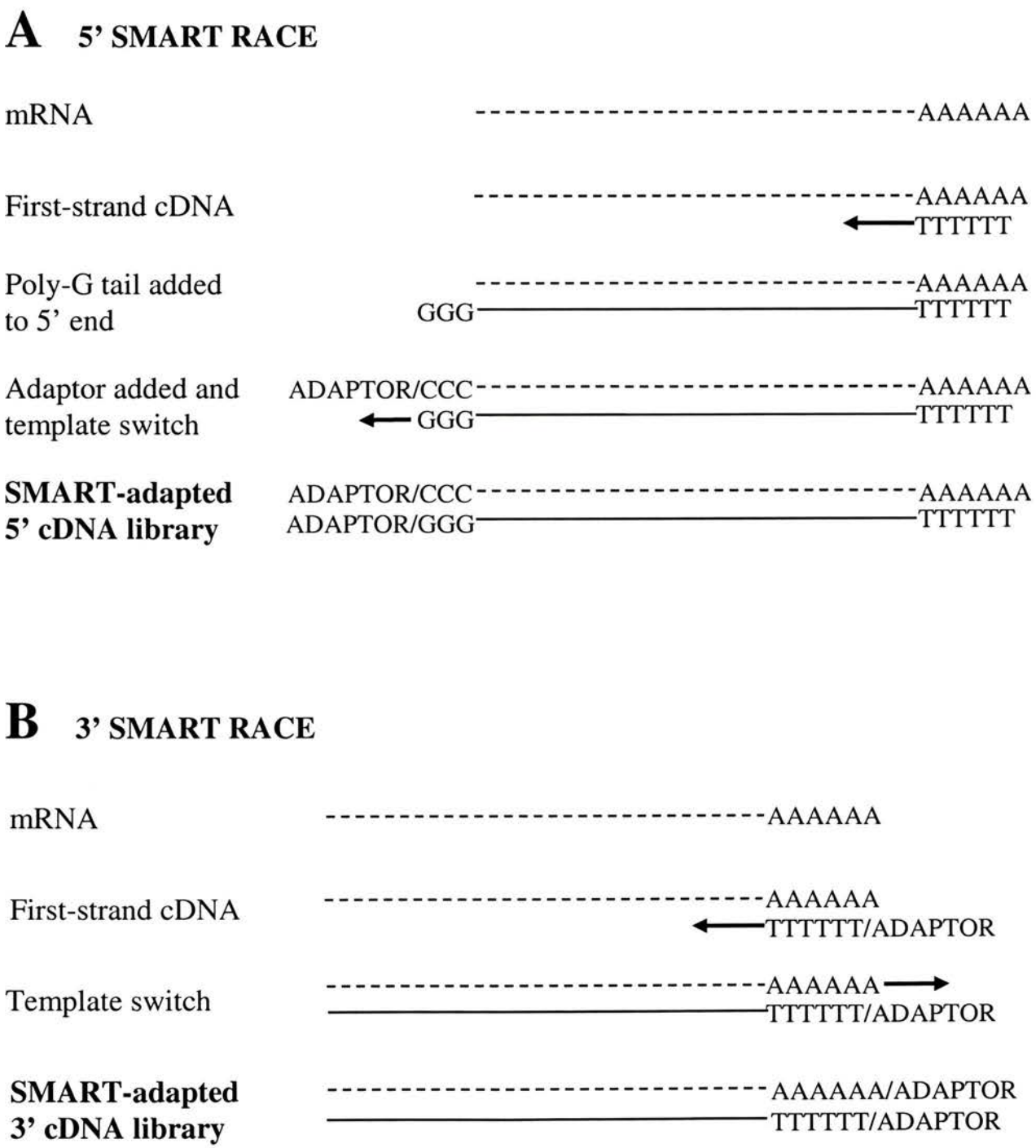
To amplify target sequence, a gene specific primer was used together with a universal primer, which annealed to the SMART adapter sequence. These were used in a 50µl PCR reaction containing 2.5µl of the diluted RACE-ready cDNA in 1xPCR buffer, 2mM dNTPs, 10pmol gene-specific primer, 2.4pmol universal primer and 1µl *Advantage 2*™ Polymerase. The PCR cycle comprised a touchdown cycle to maximise specificity. This consisted of an initial denaturation for 3 minutes at 94°C; 5 cycles of denaturation for 30 seconds at 94°C followed by a combined annealing and extension phase at 72°C; 5 cycles of denaturation for 30 seconds at 94°C, an annealing phase at 70°C and an extension phase at 72°C; 30 cycles of denaturation at 94°C for 30 seconds, annealing at 68°C and extension at 72°C. This was followed by a final extension phase for 10 minutes at 72°C. The duration of the annealing and extension phases was optimised for each target gene.

In order to generate sufficient PCR product and increase target specificity, a nested PCR reaction was performed. This used 5µl of a 1:50 dilution of the first round product in a 50µl reaction containing 10pmol of a nested gene-specific primer and 10pmol of a nested universal primer. The other components of the PCR reaction remained the same as in the first round. The PCR programme comprised an initial denaturation step of 3 minutes at 94°C followed by 27 cycles, each consisting of a 30

Figure 2.1 – SMART RACE Protocol Overview (BD Biosciences Clontech)

A 5' SMART RACE: The first strand cDNA is synthesised using a standard oligo-$dT_{(25)}$ primer, which anneals to the mRNA poly(A) tail. When the reverse transcriptase enzyme reaches the 5' end of the mRNA transcript it adds a series of guanine bases to the cDNA. The SMART oligonucleotide adaptor has a poly(C) tail, which anneals to these guanine bases. Template switch occurs and the cDNA strand is elongated to include the sequence complementary to the adaptor. A forward primer containing the adaptor sequence and a reverse gene specific primer can then be used to amplify the 5' region of the gene by PCR.

B 3' SMART RACE: First strand cDNA is synthesised using an oligo-$dT_{(25)}$ primer which already has the oligonucleotide adaptor sequence attached to it. A forward gene specific primer and a reverse primer complementary to the adaptor sequence can then be used to amplify the 3' region of the gene by PCR.


A 5' SMART RACE
mRNA
AAAAAA
First-strand cDNA
AAAAAA
TTTTTT
Poly-G tail added to 5' end
AAAAAA
GGG
TTTTTT
Adaptor added and template switch
ADAPTOR/CCC
AAAAAA
GGG
TTTTTT
SMART-adapted 5' cDNA library
ADAPTOR/CCC
AAAAAA
ADAPTOR/GGG
TTTTTT
B 3' SMART RACE
mRNA
AAAAAA
First-strand cDNA
AAAAAA
TTTTTT/ADAPTOR
Template switch
AAAAAA
TTTTTT/ADAPTOR
SMART-adapted 3' cDNA library
AAAAAA/ADAPTOR
TTTTTT/ADAPTOR

second denaturation stage at 94°C, a 30 second annealing stage at 68°C and an extension stage at 72°C. This was followed by a final extension stage for 10 minutes at 72°C. The duration of the cycling extension stage varied according to the length of the target transcript.

Each 50µl second round PCR product was mixed with 6µl loading buffer and run alongside a DNA molecular weight marker on a 1% agarose gel containing 2µg/µl EtBr. Products were excised from the gel using a sterile scalpel blade and retained for gel-extraction.

2.3.6 Gel Extractions

PCR products were extracted from the excised pieces of agarose gel using the QIAquick™ Gel Extraction Kit (Qiagen) according to the manufacturer's standard protocol. However, the initial solubilised gel slice was passed through the Qiagen column twice (instead of the standard once) to ensure maximum recovery of PCR product.

2.3.7 Cloning PCR Products into a Sequencing Vector

Gel-extracted PCR products were cloned using the pGEM™-T Easy Vector System (Promega).

2.3.7.1 Ligation and Transformation

PCR product (2µl) was ligated into the pGEM™-T Easy vector using *T4* DNA ligase and Rapid Ligation Buffer in a total volume of 10µl during an overnight incubation at 4°C. JM109 Competent Cells (40µl) were carefully added to 2µl of the ligation reaction in a prechilled 1.5ml microfuge tube. Samples were gently mixed and incubated on ice for 20 minutes. The cells/ligation mix was heat-shocked for 45 seconds at 42°C and returned to ice for two minutes. 950µl of SOC medium at room

temperature was added to the cells, which were then incubated for one hour at 37°C in a shaking incubator. A mixture containing 40µl of X-GAL (20mg/ml), 10µl IPTG (100mM) and 50µl SOC medium (100µl total) was spread evenly across the surface of an agar plate containing 100µg/ml ampicillin (D-α-aminobenzylpenicillin, a penicillin-related beta-lactam), using a glass spreader. This combination allowed selection of only transformed bacteria carrying ampicillin resistance from the plasmid and also enabled blue-white colour screening of colonies for those in which the *LacZ* gene had been disrupted by the presence of insert. Following this, 100µl of the incubated transformation mix was spread across the plate using a glass spreader. Plates were inverted and incubated overnight at 37°C and then stored at 4°C.

2.3.7.2 Colony Screening

To screen bacterial colonies for the presence of correct insert, PCR was performed on a crude lysate from five discrete white colonies (blue colonies contain an intact *lacZ* gene and do not contain insert). These colonies were selected and a sterile pipette-tip used to transfer a little of each into 50µl of water in a 0.2ml PCR tube. Agar plates were then returned to 4°C. The bacterial suspension heated to 99°C for ten minutes to lyse the cells and cell debris pelleted by centrifugation for ten minutes at full speed. The presence of the insert of the correct size was verified using 2.5µl of the supernatant (a crude DNA preparation) as a template in a 25µl PCR reaction: 1xPCR buffer containing a 2mM dNTPs, 2mM $MgCl_2$, 20pmol of vector primers T7 (5'-TAA TAC GAC TCA CTA TAG GG-3') and SP6 (5'-ATT TAG GTG ACA CTA TAG AA-3') and 0.7U *High Fidelity* DNA polymerase (Roche Diagnostics UK). The PCR programme comprised an initial denaturation step of three minutes at 94°C followed by 35 cycles, each consisting of a 30 second denaturation stage at 94°C, a one minute annealing stage at 55°C and a two minute extension stage at 72°C. This was followed by a final extension stage for ten minutes at 72°C. Each PCR product was mixed with 3µl loading buffer and run on a 1% agarose gel containing 2µg/µl EtBr.

2.3.7.3 Colony Selection

The colony displaying the strongest, clearest, most discrete product of the correct size was identified. A second sample of the selected colony was harvested from the plate using a sterile pipette-tip as before, added to 1ml of LB broth containing ampicillin (100μg/ml) and incubated for six hours at 37°C in a shaking incubator. 100μl of this day culture was transferred into 10ml of LB broth containing ampicillin (100μg/ml) and incubated overnight at 37°C in the shaking incubator.

2.3.8 Plasmid Minipreps and Maxipreps

2.3.8.1 Minipreps

Plasmid DNA was extracted from 3ml of overnight culture using the QIAprep™ Miniprep kit (Qiagen) according to the manufacturer's standard protocol. This comprised a lysis stage followed by precipitation of the cell membranes and attached bacterial DNA. The supernatant containing plasmid DNA was purified and eluted directly from the column membrane in 50μl elution buffer (TE)

2.3.8.2 Maxipreps

Plasmid DNA was extracted from 100ml of overnight culture using the Qiagen Plasmid Maxi Kit (Qiagen) according to the manufacturer's standard protocol. Unlike Minipreps, following elution from the column membrane, plasmid DNA was precipitated in isopropanol, pelleted by centrifugation and washed in 70% ethanol, before being redissolved in 200μl 1xTE.

2.3.9 Sequencing of Plasmid Clones

Sequencing reactions were carried out using 50ng extracted plasmid as template together with 1.6pmol sequencing primer (SP6 or T7), 2μl BigDye™ v3.1

Sequencing Buffer and 4µl BigDye™ Terminator v3.1 Ready Reaction Premix (Applied Biosystems); the reaction volume was made up to a total 10µl with milliQ water. The sequencing reaction consisted of an initial denaturation step for one minute at 96°C, followed by 25 cycles of ten seconds at 96°C, five seconds at 50°C and four minutes at 60°C. Samples were then stored at 4°C until ready to send for sequencing. Samples were purified and sequenced at the DNA Sequencing Service at the Institute of Cell, Animal and Population Biology (ICAPB) at the University of Edinburgh. All clones used for comparative purposes were sequenced in both directions, to give 100% overlap of the template and to ensure accuracy of the read.

2.3.10 Sequence Data Manipulation

All sequence results files were processed and contigs assembled using the Staden Package version 2002.0 (Staden *et al.* 1999; Staden *et al.* 2002). Sequences were aligned using the EMMA program (interface for ClustalW) within the Human Genome Mapping Project (HGMP) Jemboss package (Carver and Mullan 2002). The resulting multiple sequence alignment was exported into the Molecular Evolutionary Genetics Analysis program (MEGA2.1) (Kumar *et al.* 2001) to produce the figures used within this thesis.

2.4 Molecular Biology Techniques: Creating Expression and Reporter Plasmids

This section describes techniques employed to develop plasmid expression and reporter constructs.

2.4.1 Bulk-up Plasmid

For each plasmid requiring amplification, 0.5µg was used to transform 100µl XL1-Blue Subcloning Grade Competent Cells (Stratagene). These were incubated on ice for 30 minutes, heat-shocked for 25 seconds at 37°C and then incubated on ice for a

further two minutes. 950µl of L-Broth (LB) was added to each transformation and incubated for one hour at 37°C in a shaker. 100µl of each transformation reaction was plated onto LB-agar plates containing either 100µg/ml ampicillin or 50µg/ml zeocin, depending on the selectable marker present. These were allowed to grow at 37°C overnight and a single colony from each plate was picked using a sterile pipette tip and used to inoculate 1ml LB containing antibiotic at the same concentration as in the agar plates. After four hours in the shaker this starter-culture was transferred into 100ml LB containing antibiotic and grown up overnight. The following day, the culture was split into two 50ml falcon tubes and centrifuged at 1980 x g (3000rpm) for 30 minutes to pellet the bacteria. The supernatant was removed, the two pellets pooled and the plasmid DNA extracted using the Plasmid Maxi Kit (Qiagen), as described previously.

2.4.2 Restriction Digest of PCR Products with Primer-Introduced Restriction Sites

The 30µl PCR product eluted from gel-extraction was digested for three hours at 37°C using 1µl of each restriction enzyme in 1xSuRE/Cut Buffer (Roche Diagnostics). This digestion was cleaned up using QIAquick™ PCR Purification Kit (Qiagen) and eluted into 30µl. The appropriate SuRE/Cut Buffer was selected for each combination of restriction enzymes used.

2.4.3 Restriction Digest and Dephosphorylation of Plasmid Open Ends

Approximately 5µg of plasmid was digested for two hours at 37°C using 1µl of each restriction enzyme and 1xSuRE/Cut Buffer in a total volume of 20µl (Roche Diagnostics). This digestion was cleaned up using QIAquick™ PCR Purification Kit (Qiagen) and eluted into 30µl. The appropriate SuRE/Cut Buffer was selected for each combination of restriction enzymes used.

The open ends of the digested plasmid were dephosphorylated using shrimp alkaline phosphatase (SAP) for one hour at 37°C (Roche Diagnostics). The plasmid was then

cleaned once again using the QIAquick™ PCR Purification Kit. Dephosphorylation of the open ends prevents the plasmid from ligating back together during the later ligation process, favouring insertion of the PCR product.

2.4.4 Ligations

The Rapid DNA Ligation Kit (Roche Diagnostics) was used to insert a digested PCR product into a digested, dephosphorylated plasmid at a molar ratio of 10:1 insert to plasmid. 8µl of the 20µl ligation reaction was used to transform 100µl XL1-Blue Subcloning Grade Competent Cells (Stratagene), which were incubated on ice for 30 minutes and then heat-shocked for 25 seconds at 37°C. After a further two minutes on ice, 300µl of LB was added to the transformation reaction and incubated for 1 hour at 37°C in a shaker. Two 150µl aliquots of this culture were plated out onto two separate LB-agar plates containing either 100µg/ml ampicillin or 50µg/ml zeocin as appropriate. The following day, five colonies from each plate were picked and used to inoculate 10ml volumes of LB containing the appropriate antibiotic. Cultures were incubated overnight and plasmid DNA extracted from 3ml of the culture using QIAprep™ Miniprep Kit (Qiagen) as previously described. To confirm that ligations had occurred as expected, 1µl of extracted plasmid was digested and run on a 1% agarose gel to confirm its banding pattern. Once the correct plasmid had been obtained and identified, 2ml of the 10ml overnight culture was added to 100ml LB containing antibiotic and incubated overnight. Plasmid DNA was extracted as previously described using the QIAGEN Plasmid Maxi Kit.

2.5 Molecular Biology Techniques: Tissue Culture and Reporter Assays

This section describes cell lines and techniques utilised during tissue culture. It also includes details on processing of cells for reporter assays.

2.5.1 Cell Lines

2.5.1.1 Hepa 1 Cells

Hepa 1 cells (murine hepatoma line) are derived from BW7756 tumour in a C57L mouse. They are easy to maintain and are frequently used for pilot studies. These cells were kindly provided by Dr. Bruce Whitelaw (Roslin Institute).

2.5.1.2 Vero Cells

Vero cells are derived from the kidney of an adult African green monkey (*Cercopithecus aethiops*) and were purchased from the European Collection of Cell Cultures. Vero cells have been previously used for *in vitro* African swine fever virus infection studies (Miskin *et al.* 2000; Alonso *et al.* 2001). By using them in this study, we provide the flexibility for future infection studies.

2.5.1.3 Murine Embryonic Fibroblasts (MEFs)

Murine embryonic fibroblasts (MEFs) (wild type p65 +/+ and p65 -/-) were kindly donated by Professor Ronald Hay at the University of St. Andrews. To obtain the p65 -/- cells, two p65 +/- mice were bred together and MEF cells derived from embryos which inherited a p65 -/- genotype. These embryos would have otherwise died at day 15-16 *in utero* from the lethal knockout.

2.5.2 Cell Maintenance

2.5.2.1 Culturing Cells

All cell lines used were grown in T-75 tissue culture flasks in a 'plating volume' of 15ml Dulbecco's Modified Eagle Medium (DMEM) containing 10% foetal calf serum (FCS) and 2mM L-glutamine (LG). Cells were cultured in a humidified 37°C, 5% carbon dioxide (CO_2) incubator. While being maintained, cells were passaged

twice weekly as they reached 70-80% confluence. The old media was removed and the cells washed with 10ml phosphate-buffered saline (PBS, pH7.4). All cell lines were adherent and digested from the surface of the flask using 1.5ml 1x trypsin-EDTA (TE); the flask was tapped vigorously after a few minutes to help dislodge cells. After five minutes, 8.5ml fresh media, containing FCS, was added to neutralise the trypsin. A 5ml pipette was used to break up any cell clumps by pipetting up and down, and 1.5ml of cell suspension transferred into a new flask containing 13.5ml fresh media.

2.5.2.2 Freezing Cells

Cells were initially suspended as per the passage protocol described in Section 2.4.2.1. Once fresh medium had been added to neutralise the trypsin, the suspension was transferred into a 15ml tube and centrifuged for five minutes at 220 x g (1000 rpm) to pellet the cells. The media was removed and replaced with 5ml freezing media (growth media with FCS and LG, plus an additional 10% DMSO). The pellet was resuspended and divided into five 1ml volumes in 1.5ml cryotubes. These were then placed in a -80°C freezer in a polystyrene box and frozen at approximately 1°C per minute before being transferred into metal racks for long-term storage 24 hours later.

2.5.3 Transfection of Plasmid DNA into Cultured Cells

Cells were cultured such that they were at high density (80-90% confluent) on the day of transfection. In preparation for transfection, the old media was removed, cells washed with PBS and 12ml unsupplemented DMEM added (without FCS or LG). For each T-75 flask, 30µg of plasmid DNA was required for transfection. This was mixed with 1.5ml of unsupplemented DMEM in a 4ml tube. In a separate tube, 70µl Lipofectamine 2000 Reagent (Invitrogen) was added to 1.5ml unsupplemented DMEM. These two solutions were then mixed together to give a total volume of 3ml. This transfection mixture was incubated for 20 minutes at room temperature to allow

the lipid carrier and plasmid to bind. This transfection mix was then added to the media in the flask, bringing the plating volume up to 15ml. After five hours in the 37°C, 5% CO_2 incubator, the transfection media was removed, the cells washed with PBS and the media replaced with normal DMEM growth media containing 10% FCS and 2mM LG.

2.5.3.1 Selection of Stably Transfected Cells

For selection of stable cells in which the plasmid DNA has successfully integrated into the host-cell nuclear DNA, antibiotic (Zeocin or G418) was added to the media 24 hours post transfection. The concentration and type of antibiotic used depended on the cell type and plasmid backbone used (Table 2.3). In order to determine the concentration required for selection, cells were treated with a range of antibiotic concentrations (0, 200, 400, 600, 800, 1000μg/ml) prior to transfection. This established the approximate concentration of antibiotic required to kill 100% of the cells within 10 days. Using the results of the first assay, a second more-focussed experiment was undertaken. For example, following an initial estimate of optimal kill at 600μg/ml, the antibiotic range used for the second assay might include: 450, 500, 550, 600, 650, 700μg/ml.

Zeocin (Cayla) is a commercial preparation containing Phleomycin D1, a member of the blebomycin family of antibiotics. It acts primarily by binding to, and degrading, DNA. G418 (Sigma Aldrich) is synthetic aminoglycoside antibiotic similar to neomycin and gentamycin. Aminoglycosides act by interfering with ribosomal function and blocking protein synthesis/elongation.

Transfected cell cultures were washed with PBS and the selection media replaced every other day for 10-14 days until colonies originating from successfully transfected cells were clearly visible. The cells were trypsinised and pooled to normalise for different insertion sites. Since this population can change over time (as cells with different insertion sites may display growth advantage), twelve colonies from a second transfection carried out in a 10cm culture dish were isolated using

Table 2.3

Type and concentration of antibiotic required for selection of stably transfected cells. The concentration required was dependent on cell type, established by the concentration needed to kill 100% untransfected cells within 10 days.

Cell type	Plasmid backbone	Antibiotic	Concentration (μg/ml)
Hepa 1	pBudCE4.1	Zeocin	200
Vero	pFLAG	G418	600
MEFs p65 -/-	pFLAG	G418	1200-1600
MEFs p65 +/+	pFLAG	G418	100

glass rings, held in place with sterile grease. 150µl 1xTE was added to the centre of each ring and left for five minutes to digest the colony from the dish. The TE was then pipetted up and down to ensure the colony was in suspension and transferred to 1ml selection medium in individual wells in a 24-well plate. The collected cells were cultured for 48 hours, then trypsinised and transferred into 6-well plates and cultured for a further 48 hours. These cells were then frozen down and stored at -80°C. Each culture represents a clonal population which displays stable characteristics and can be used at a later date for more in-depth analysis.

2.5.4 Luciferase Reporter Assays

To measure luciferase levels in cells transfected with a double-reporter plasmid, the Dual-Reporter Assay System (Promega) was used. Media was removed from the cells in a six or twelve-well plate (containing ~$1x10^6$ and ~$3.5x10^5$ cells per well respectively), which were then washed with 1ml PBS. Cells were lysed *in situ* using 0.5ml passive lysis buffer (from Promega Dual-Reporter Assay kit) and placed on a rocking platform for 15 minutes. Lysates were removed into 1.5ml microfuge tubes, pipetted up and down to disrupt cell clumps and centrifuged for one minute at full speed to pellet any cell debris. The supernatant was then removed to a clean microfuge tube. To assay for luciferase activity, three 20µl volumes of lysate from each sample were added to wells in an opaque 96-well microtitre plate. The BioLumat luminometer was set up to inject 100µl of Luciferase Assay Reagent II (LAR II) into each well, wait for two seconds and then read the output from the firefly luciferase for ten seconds. The luminometer then injected 100µl Stop & Glo Reagent, which halted the firefly luciferase reaction, waited for two seconds and then read renilla luciferase activity for ten seconds.

2.5.5 Total Protein Assay

In order to normalise for cell number, total protein assays were also carried out on cell lysates using the BCA (bicinchoninic acid) Protein Assay Kit (Pierce). Firstly,

nine standard solutions were made using bovine serum albumin (BSA), from which a standard curve could be derived. Standard solutions were made up in the same diluent (1x passive lysis buffer) as the experimental samples. Three 25µl replicates of each standard and experimental sample were added to wells of a clear 96-well microtitre plate, followed by 200µl of BCA assay solution. The plate wells were sealed with an adhesive cover, samples mixed thoroughly and incubated for 30 minutes at 37°C. Measurements of protein content were then taken using a colorimetric plate reader at a wavelength of 590nm. This assay is based on the biuret reaction, where protein reduces Cu^{2+} to Cu^{+} in an alkaline environment. The purple/blue colour used in the assay is formed by the chelation of two BCA molecules with a Cu^{+} ion.

The protein standard results were used to plot a graph and the equation of the line of best fit calculated in Microsoft Excel. This equation was then used to determine the protein concentration of each experimental well and an average protein content taken from the three replicates of each sample.

2.6 Molecular Biology Techniques: Protein Extraction and Analysis

Based on the results of the comparative sequence analysis described in Chapter 3, it was decided to undertake a pilot study involving overexpression of p65 *in vitro* (Chapter 5). This study required protein extraction and analysis, the protocols for which are detailed below.

2.6.1 Protein Extraction

For protein extraction, cells in culture were washed in PBS, trypsinised from the culture flask and centrifuged for five minutes at 220 x g (1000rpm). All media was removed and the pellet washed with PBS and spun down again. Once this wash step had been repeated, the pellet was resuspended in 100µl homogenisation buffer containing antiproteases (for constituents see Table 2.2) and pipetted up and down to lyse and homogenise the cells. The homogenate was incubated on ice for 30 minutes

and centrifuged for one minute at 16000 x g (13000rpm) to pellet the cell debris. The protein-containing supernatant was removed to a clean microfuge tube. A 1:50 and a 1:100 dilution were made up in distilled water to assay for total protein concentration using the BCA kit (Pierce) as previously described (Section 2.4.5).

2.6.2 Western Blotting

Western blotting is a technique used to label and visualise a specific protein. Firstly, all proteins in a sample are separated, primarily by molecular weight, using electrophoresis through a polyacrylamide gel. The proteins are then transferred onto a nitrocellulose membrane and the target protein labelled using a specific (primary) antibody. A tagged secondary antibody (targeted against the primary antibody) is then utilised to amplify the signal and enable visualisation of the hybridised complex, achieved using chemiluminescence.

2.6.2.1 Running Extracted Protein

Based on the results of the total protein assay, 50μg protein was added to 10μl 4x Sample Buffer (Invitrogen) with 2μl β-mercaptoethanol (reducing agent) and made up to 40μl with distilled water. The sample was heated at 70°C for ten minutes to denature the proteins and loaded onto a NuPAGE Bis-Tris sodium dodecylsulphate polyacrylamide gel electrophoresis (SDS-PAGE) mini-gel (Invitrogen) using a Hamilton syringe. A protein molecular weight marker lane was also included and two gels run simultaneously for 1 hour at 200V using an XCell *SureLock* Mini-Cell (Invitrogen) system. The outer chamber was filled with ~700ml NuPAGE MOPS SDS Running Buffer (Invitrogen) and the inner chamber filled with 200ml of the same running buffer, but containing 0.5ml NuPAGE Antioxidant (Invitrogen).

2.6.2.2 Transferring Proteins onto a Nitrocellulose Membrane

In order to transfer the proteins run on the gel onto nitrocellulose membrane, the XCell II Blot Module (Invitrogen) was assembled, as detailed in the manufacturer's protocol, using blotting pads, filter paper and two pieces of 0.2μm nitrocellulose membrane (all presoaked in transfer buffer – Table 2.2) and two gels. The layers were 'rolled' at various stages with the barrel of a glass pipette to push out any air bubbles introduced during the assembly process. The module was locked shut and placed back into the empty XCell *SureLock* tank. The inner chamber was filled with transfer buffer and the outer chamber filled with water to keep the apparatus cool. The transfer was carried out at 30V for 1 hour.

2.6.2.3 Visualising Total Protein

Once the proteins had been transferred onto nitrocellulose, the blot module was disassembled and the nitrocellulose membranes placed in a container of Ponceau S stain on a shaker. All protein stained red and the key marker bands were drawn on with a permanent marker pen. The blot was temporarily placed between acetate sheets to be photocopied. The blot was then trimmed and returned to water on a shaker to wash off the Ponceau S.

2.6.2.4 Blocking the Blot

In order to prevent any non-specific binding by the primary antibody, the blot was blocked overnight at 4°C with shaking in a 4% solution of non-fat dried milk (NFDM) in Tris-buffered saline (TBS).

2.6.2.5 Primary Antibody Hybridisation

The blot was removed from the overnight blocking solution and washed three times for five minutes in 5% NFDM in TBST (TBS containing 0.1% Tween 20). The

required antibody was diluted to the appropriate concentration in 5% NFDM/TBST and left to incubate on the shaker at room temperature (see relevant experiment for incubation time).

After incubation, the antibody-containing 5% NFDM/TBST solution was removed and the blot washed three times for ten minutes each in fresh 5% NFDM/TBST with shaking.

2.6.2.6 Secondary Antibody Hybridisation

The secondary antibody was supplied as a conjugate with horse radish peroxidase (HRP) (Abcam). This was used at a concentration of 1:5000 (2µl in 10ml) in 5% NFDM/TBST and hybridised to the blot for one hour with shaking. The blot was then washed to remove unbound antibody: four times for five minutes each in 5% NFDM/TBST, three times for five minutes in TBST (no NFDM) and twice for five minutes in TBS (no Tween 20 or NFDM).

2.6.2.7 Detection of Horse Radish Peroxidase (HRP)

The SuperSignal™ West Pico Chemiluminescent Substrate kit (Pierce) was used to detect the HRP-conjugated secondary antibody. Equal volumes of the stable peroxidase and enhancer solutions were mixed and applied to the blot (enough to just cover the surface) for five minutes. The blot was then placed between two acetate sheets and sealed in with tape before being exposed to autoradiograph film in a darkroom for periods of approximately 20 seconds, one minute and two minutes to provide a range of exposures. The exposed film was developed in an automatic developer.

2.7 Statistical Analyses

Professional assistance with all statistical analyses described was kindly given by Anthea Springbett (Roslin Institute). The computer package used was Genstat (7th edition (2003), Lawes Agricultural Trust).

All data were examined for evidence of non-Normality before analysis. In all cases the data were found to be suitable for analysis on the original scale and no transformations were performed.

The data were normalised using either renilla luciferase activity (TNFα promoter study, chapter 4) or total protein concentration (p65 study, chapter 5). For the comparative TNFα promoter activity study, dose response curves were fitted separately using regression analysis for each species and treatment. Both linear and curvilinear models were fitted and all parameters assessed by significance tests (F-tests and/or t-tests). The simplest models providing a good fit to the data were chosen. A small number of data points were omitted because they appeared to be outliers.

In the comparative p65 study, the effects of p65 genotype and treatment conditions on firefly luciferase activity were assessed by Analysis of Variance (after normalisation). The interaction between treatments was fitted and analysed for significance by F-test. In the absence of a significant interaction only the treatment effects are reported.

3 Results: Comparative Sequencing Data

In vitro evidence suggests that the ASFV-encoded protein, A238L, modulates the host immune response by interacting with key elements of the porcine NFκB and NFAT signalling pathways (Powell *et al.* 1996; Miskin *et al.* 1998; Miskin *et al.* 2000; Tait *et al.* 2000). ASFV also expresses p54 on the surface of its capsid, which interacts with the cargo-specific tether, light chain dynein. This enables the intracellular transport of virus particles by cytoplasmic dynein (Rodriguez *et al.* 1996; Alonso *et al.* 2001). It was therefore decided to sequence and compare six proteins targeted or mimicked by ASFV (light chain dynein, cyclophilin A, IκBα, the NFAT2 regulatory domain, calcineurin A and p65) in the ASFV-resistant warthog, susceptible domestic pig and phenotypically unknown babirusa. Furthermore, the proinflammatory cytokine, TNFα, is massively upregulated during ASFV infection of the domestic pig (Gómez-Villamandos *et al.* 1996; Gómez del Moral *et al.* 1999; Carrasco *et al.* 2002). As TNFα expression is primarily driven by a *cis*-acting promoter, a 1.6kbp region of sequence located immediately upstream of the TNFα start codon was sequenced and compared in the three species listed above. It was hypothesised that sequence variation between the ASFV-mimicked/targeted proteins and the TNFα promoter may determine the different pathological outcomes of ASFV infection in suid species of varying susceptibility.

Reverse transcription (RT)-PCR reactions were performed to amplify light chain dynein, cyclophilin A, IκBα, the NFAT2 regulatory domain, calcineurin A and p65 from mRNA isolated from an ASFV-susceptible domestic pig, a resistant warthog and a phenotypically unknown babirusa. The TNFα promoter was also amplified by PCR from genomic DNA isolated from the same animals. All amplicons were cloned into a sequencing vector and plasmids containing the sequence of interest were subsequently isolated. Gene-specific primer sequences are given in Table 3.1 and the optimised PCR cycling conditions for each gene are outlined in Table 3.2. These sequences were aligned together with human sequence obtained from the EMBL database.

Table 3.1

Sequences of all gene-specific PCR primers used to amplify targets for comparative sequence analysis. The following ambiguity codes are also used: Y = C or T and K = T or G.

Gene	Forward Primer (5'-3')	Reverse Primer (5'-3')
Light chain dynein	CYA GCA CCT YCC CYA GGA G	ACT GGG TGT KTG GCA CAG TC
Cyclophilin A	ATC TNT CAG TGC TGC TCA GC	CAG AAG GAA TGG TCT GAT GG
IkBa (cDNA)	AAG GAG CGG CTA CTG GAT G	CAT GGT CTT TTA GAC ACT TTC CA
IkBa (5' DNA)	CTC ATC GCA GGG AGT TTC TC	TCC TCG TCC TTC ATG GAG TC
NFAT2 (regulatory)	ATC TCA GCT GTT GGG TCA GC	AGT GAG GGT GAG TGG TCC AG
Calcineurin Aβ	CCC AAC ACA TCG TTT GAC AT	ATG TGA GAG TCC CTG GGA AG
p65	GAC CTC TTC CCC CTC ATC TT	CCC CTT AGG AGC TGA TCT GA
TNFα promoter	GAT AAA GTG ACA GAA GGC ATG G	ATA ACC TCG AAG TGC AGT AGG C

Table 3.2

Optimised cycling conditions for the seven PCR reactions used to amplify targets for comparative sequence analysis. Light chain dynein, Cyclophilin A, IκBα, the NFAT2 regulatory domain, calcineurin A and p65 used a standard 35-stage cycle, where as the TNFα promoter required a 36-cycle touchdown reaction.

		Light chain dynein		*Cyclophilin A*		*IκBα* (cDNA)		*IκBα* (5' DNA)		*NFAT2 (regulatory)*		*Calcineurin Aβ*		*p65*	
Initial Denaturation		94°C 3 minutes													
Denaturation	x35 cycles	94°C	30 sec	94°C	30 sec	94°C	30 sec	94°C	60 sec	94°C	30 sec	94°C	60 sec	94°C	30 sec
Annealing		55°C	30 sec	55°C	30 sec	55°C	30 sec	55°C	60 sec	55°C	30 sec	57°C	60 sec	55°C	30 sec
Extension		72°C	60 sec	72°C	60 sec	72°C	60 sec	72°C	60 sec	72°C	60 sec	72°C	90 sec	72°C	105 sec
Final Extension		72°C 10 minutes													

		TNFα promoter	
Initial Denaturation		95°C 12 minutes	
Denaturation	x2 cycles	94°C	45 sec
Annealing		63°C	60 sec
Extension		72°C	120 sec
Denaturation	x2 cycles	94°C	45 sec
Annealing		62°C	60 sec
Extension		72°C	120 sec
Denaturation	x2 cycles	94°C	45 sec
Annealing		61°C	60 sec
Extension		72°C	120 sec
Denaturation	x30 cycles	94°C	45 sec
Annealing		59°C	60 sec
Extension		72°C	120 sec
Final Extension		72°C 10 minutes	

3.1 Light Chain Dynein

In suid species considered here, light chain dynein (LCD) nucleotide sequence is 270bp in length and encodes a primary protein of 90 amino acids (Figure 3.1). The entire nucleotide coding sequence is completely conserved between all three species.

The domestic pig light chain dynein sequence, as published by Alonso *et al.* (2001), was significantly different to all other mammalian LCD sequences. This uniqueness naturally identified it as a gene worthy of further study in the context of this work (Figure 3.2). However, as described above, this thesis has shown LCD to be conserved in all suid species. As a result, Dr. Alonso was contacted to discuss this discrepancy. She confirmed that an error had been made in their sequence, which has since been amended and now resembles that of the other suids (June 2003).

3.2 Cyclophilin A

In suid species, cyclophilin A nucleotide sequence is 495bp in length and encodes a primary protein of 165 amino acids. In humans, the sequence is 498bp in length, encoding a primary protein comprising 166 amino acids (Figure 3.3). The additional amino acid is a glutamic acid residue located at the carboxyl terminus. In addition to these three bases at the 3' end, the human sequence differs from the domestic pig at 28 nucleotides. Only two of these differences encode alternative amino acids, glutamic acid (as opposed to aspartic acid) at position 84 and leucine (as opposed to isoleucine) at position 164.

Four nucleotide differences were identified between the three suid species examined (Table 3.3a). All are silent between the domestic pig and warthog. However, the babirusa has two mutations at the amino acid level: an alanine (as opposed to threonine) at position 107 and a serine (as opposed to alanine) at position 141 (Table 3.3b).

Figure 3.1 – *Light chain dynein* Comparative Sequence Analysis

SS	*Sus scrofa*	Domestic Pig
SS1	*Sus scrofa* (BF189957)	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
BB	*Babyrousa babyrussa*	Babirusa
SSX	*Sus scrofa* (AF436777)	Domestic Pig (Alonso *et al.* 2001)

A *Light chain dynein* (*LCD*) complete nucleotide coding sequence. *LCD* coding sequence is 270 bases in length and is completely conserved between all suid species represented here. The domestic pig sequence (SSX) published by Alonso *et al.* (2001) contained a number of mistakes, which have since been corrected.

B The predicted primary protein sequence for light chain dynein is 90 amino acids in length.

A

```
SS  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [ 60]
SS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [ 60]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [ 60]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [ 60]
SSX ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [ 60]

SS  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [120]
SS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [120]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [120]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [120]
SSX ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... [120]

SS  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A [180]
SS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [180]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [180]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [180]
SSX ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ... ... .A. [180]

SS  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [240]
SS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [240]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [240]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... [240]
SSX .CG ... ... ... ... ... ... ... ... ... ... ... ... ... G.T GC- --- --- --- --- --- --- [240]

SS  ... ... ... ... ... ... ... ... ... ... ...  [270]
SS1 ... ... ... ... ... ... ... ... ... ... ...  [270]
BB  ... ... ... ... ... ... ... ... ... ... ...  [270]
PA  ... ... ... ... ... ... ... ... ... ... ...  [270]
SSX --- --- --A ..T ... ... ... ... ... ... ...  [270]
```

B

```
SS  MCDRKAVIKN ADMSEEMQQD SVECATQALE KYNIEKDIAA HIKKEFDKKY NPTWHCIVGR  [60]
SS1 .......... .......... .......... .......... .......... ..........  [60]
BB  .......... .......... .......... .......... .......... ..........  [60]
PA  .......... .......... .......... .......... .......... ..........  [60]
SSX .......... .......... .......... ..K....... .......... .........K  [60]

SS  NFGSYVTHET KHFIYFYIGQ VAILLFKSG*  [90]
SS1 .......... .......... ..........  [90]
BB  .......... .......... ..........  [90]
PA  .......... .......... ..........  [90]
SSX T......... ..V------- ---.......  [90]
```

Figure 3.2 – Mammalian *Light chain dynein* Alignment

HS	*Homo sapiens*	U32944	Human
MF	*Macaca fascicularis*	AB056397	Macaque
MM	*Mus musculus*	BC034258	Mouse
RN	*Rattus norvegicus*	U66461	Rat
OC	*Oryctolagus cuniculus*	AF008304	Rabbit
SSX	*Sus scrofa*	AF436777	Domestic Pig (Alonso *et al.* 2001)

Light chain dynein (*LCD*) nucleotide coding sequence is highly conserved across mammals and the primary protein sequence is entirely conserved. The domestic pig sequence (SSX) published by Alonso *et al.* (2001) contained several mistakes, which have since been corrected.

A

```
HS   ATG TGC GAC CGA AAG GCC GTG ATC AAA AAT GCG GAC ATG TCG GAA GAG ATG CAA CAG GAC  [ 60]
MF   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 60]
MM   ... ... ... ..G ... ..G ... ... ... ... ..A ... ... ... ... ... ... ... ... ...  [ 60]
RN   ... ... ... ..G ... ..G ... ... ... ... ..A ... ... ... ... ... ... ... ... ...  [ 60]
OC   ... ... ... ..C ... ..G ... ... ..G ..C ... ... ... ... ... ... ... ... ... ...  [ 60]
SSX  ... ..T ... ... ... ... ... ... ... ... ..C ..T ... ... ..G ... ... ... ... ..T  [ 60]

HS   TCG GTG GAG TGC GCT ACT CAG GCG CTG GAG AAA TAC AAC ATA GAG AAG GAC ATT GCG GCT  [120]
MF   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [120]
MM   ... ... ... ... ... ..C ... ... T.. ... ..G ... ... ..C ... ... ..T ... ... ..C  [120]
RN   ... ... ... ... ... ... ... ... T.. ... ..G ... ... ... ... ... ..T ..C ... ..C  [120]
OC   ... ... ... ... ..C ... ..A ... ... ... ... ... ... ..C ... ... ... ..C ... ..C  [120]
SSX  ... ... ... ..T ... ... ... ..A ... ... ... ..T ..G ... ... ... ... ... ... ..C  [120]

HS   CAT ATC AAG AAG GAA TTT GAC AAG AAG TAC AAT CCC ACC TGG CAT TGC ATC GTG GGG AGG  [180]
MF   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ...  [180]
MM   ... ... ... ... ..G ... ... ... ... ... ..C ..T ... ... ..C ... ..T ... ..C C.A  [180]
RN   ... ... ... ... ..G ... ... ... ... ... ..C ... ... ... ..C ... ... ... ..C C..  [180]
OC   ..C ... ... ... ..G ... ... ... ... ... ..C ... ... ... ..C ... ..T ... ... ...  [180]
SSX  ... ... ... ... ..G ... ... ..A ... ... ..C ... ..T ... ..C ... ..T ... ... .A.  [180]

HS   AAC TTC GGT AGT TAT GTG ACA CAT GAA ACC AAA CAC TTC ATC TAC TTC TAC CTG GGC CAA  [240]
MF   ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [240]
MM   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ..G  [240]
RN   ... ... ... ..C ..C ... ... ..C ..G ... ... ... ... ... ... ... ... ... ..T ..G  [240]
OC   ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G  [240]
SSX  .CG ... ... ... ..C ... ... ... ... ... ... ... --- --- --- --- --- --- --- ---  [240]

HS   GTG GCC ATT CTT CTG TTC AAA TCT GGT TAA  [270]
MF   ... ... ... ... ... ... ... ... ... ...  [270]
MM   ... ... ... ... ... ... ... ... ... ...  [270]
RN   ... ... ... ..C ... ... ... ... ... ...  [270]
OC   ... ... ... ... ... ... ... ... ... ..G  [270]
SSX  --- .TT GCA ... ... ... ... ... ... ...  [270]
```

B

```
HS   MCDRKAVIKN ADMSEEMQQD SVECATQALE KYNIEKDIAA HIKKEFDKKY NPTWHCIVGR  [60]
MF   .......... .......... .......... .......... .......... ..........  [60]
MM   .......... .......... .......... .......... .......... ..........  [60]
RN   .......... .......... .......... .......... .......... ..........  [60]
OC   .......... .......... .......... .......... .......... ..........  [60]
SSX  .......... .......... .......... ..K....... .......... .........K  [60]

HS   NFGSYVTHET KHFIYFYLGQ VAILLFKSG*  [90]
MF   .......... .......... ..........  [90]
MM   .......... .......... ..........  [90]
RN   .......... .......... ..........  [90]
OC   .......... .......... ..........  [90]
SSX  T......... ..-------- -VA.......  [90]
```

Figure 3.3 – *Cyclophilin A* Comparative Sequence Analysis

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
BB	*Babyrousa babyrussa*	Babirusa
HS	*Homo sapiens*	Human (BC005320)

A *Cyclophilin A* complete nucleotide coding sequence. Suid *cyclophilin A* coding sequence is 495 bases long (including the stop codon). Human *cyclophilin A* is 498 base pairs in length, due to an additional codon at the 3' end.

B The predicted primary protein sequence for cyclophilin A is 165 amino acids in length in suids and 166 amino acids in humans. The protein sequence is conserved between the domestic pig and warthog, which differs from the babirusa at two residues.

A

```
SS ATG GTT AAC CCC ACC GTC TTC TTC GAC ATC GCC GTC GAT GGC GAG CCC TTG GGC CGC GTC  [ 60]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 60]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 60]
HS ... ..C ... ... ... ..G ... ... ... ..T ... ... ..C ... ... ... ... ... ... ...  [ 60]

SS TCC TTC GAG CTG TTT GCA GAC AAA GTT CCA AAG ACA GCA GAA AAC TTC CGT GCT CTG AGC  [120]
PA ... ... ... ... ... ..T ... ... ... ... ... ... ... ... ... ..T ... ... ... ...  [120]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ...  [120]
HS ... ..T ... ... ... ... ... ..G ..C ... ... ... ... ... ..T ..T ... ... ... ...  [120]

SS ACT GGG GAG AAA GGA TTT GGT TAT AAA GGT TCC TGC TTT CAC AGA ATA ATT CCA GGA TTT  [180]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [180]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [180]
HS ... ..A ... ... ... ... ... ... ..G ... ... ... ... ... ... ..T ... ... ..G ...  [180]

SS ATG TGC CAG GGT GGT GAC TTC ACA CGC CAT AAT GGC ACT GGT GGC AAG TCC ATC TAT GGA  [240]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [240]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [240]
HS ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G  [240]

SS GAG AAA TTT GAT GAT GAG AAT TTT ATC CTG AAG CAT ACG GGT CCT GGC ATC TTG TCC ATG  [300]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [300]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [300]
HS ... ... ... ..A ... ... ..C ..C ... ..A ... ... ... ... ... ... ... ... ... ...  [300]

SS GCA AAT GCT GGC CCC AAC ACA AAC GGT TCC CAG TTT TTC ATT TGC ACT GCC AAG ACT GAG  [360]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [360]
BB ... ... ... ... ... ... G.. ... ... ... ... ... ... ... ... ... ... ... ... ...  [360]
HS ... ... ... ..A ... ... ... ..T ... ... ... ... ... ..C ... ... ... ... ... ...  [360]

SS TGG TTG GAT GGC AAA CAT GTG GTC TTT GGC AAA GTG AAA GAG GGC ATG AAT ATT GTG GAA  [420]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [420]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [420]
HS ... ... ... ... ..G ... ... ..G ... ... ... ... ... ..A ... ... ... ... ... ..G  [420]

SS GCC ATG GAG CGC TTT GGG TCC AGG AAT GGC AAG ACC AGC AAG AAG ATC ACC ATT GCT GAC  [480]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [480]
BB T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [480]
HS ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [480]

SS TGT GGA CAA ATC TAA ---  [498]
PA ... ... ... ... ... ---  [498]
BB ... ... ... ... ... ---  [498]
HS ... ... ... C.. G.. TAA  [498]
```

B

```
SS MVNPTVFFDI AVDGEPLGRV SFELFADKVP KTAENFRALS TGEKGFGYKG SCFHRIIPGF MCQGGDFTRH  [ 70]
PA .......... .......... .......... .......... .......... .......... ..........  [ 70]
BB .......... .......... .......... .......... .......... .......... ..........  [ 70]
HS .......... .......... .......... .......... .......... .......... ..........  [ 70]

SS NGTGGKSIYG EKFDDENFIL KHTGPGILSM ANAGPNTNGS QFFICTAKTE WLDGKHVVFG KVKEGMNIVE  [140]
PA .......... .......... .......... .......... .......... .......... ..........  [140]
BB .......... .......... .......... ......A... .......... .......... ..........  [140]
HS .......... ...E...... .......... .......... .......... .......... ..........  [140]

SS AMERFGSRNG KTSKKITIAD CGQI*-  [166]
PA .......... .......... .....-  [166]
BB S......... .......... .....-  [166]
HS .......... .......... ...LE*  [166]
```

Table 3.3

A The four nucleotide differences existing between domestic pig, warthog and babirusa *cyclophilin A* sequence.

B The two amino acid differences between the babirusa cyclophilin A primary protein sequence and the other suids. Primary protein sequence is conserved between the domestic pig and warthog.

A

Nucleotide Position	Domestic pig	Warthog	Babirusa	Codon Position	Silent mutation? Which species differs?
78	A	T	A	3rd	Silent
108	C	T	T	3rd	Silent
319	A	A	G	1st	Babirusa
421	G	G	T	1st	Babirusa

B

Amino Acid Position	Corresponding codon	Domestic pig	Warthog	Babirusa
107	319-321	Threonine (T)	Threonine (T)	Alanine (A)
141	421-423	Alanine (A)	Alanine (A)	Serine (S)

3.3 IκBα

The majority of the IκBα coding sequence was obtained in the normal manner by cloning PCR products derived from cDNA amplification. However, because the 5' untranslated region is GC-rich, it was necessary to design the forward primer within the 5' coding sequence of the first exon. In order to achieve full coding sequence, the remaining 5' sequence was obtained following cloning of a PCR reaction directly from DNA. This was achieved using (1) a forward primer designed in the promoter region from sequence Z35483 (available in the public database) and (2) a reverse primer within the first exon, downstream of the forward primer used in the initial PCR (Figure 3.4). In suids, IκBα nucleotide sequence is 945bp in length and encodes a primary protein of 315 amino acids. In humans, the sequence is 954bp in length, encoding a primary protein of 318 amino acids (Figure 3.5). Eleven nucleotide differences were identified between the three suid species examined (Table 3.4a). All differences are silent between the domestic pig and warthog. However, the babirusa possesses one mutation at the amino acid level - a glutamic acid at position 170 (as opposed to glutamine) (Table 3.4b).

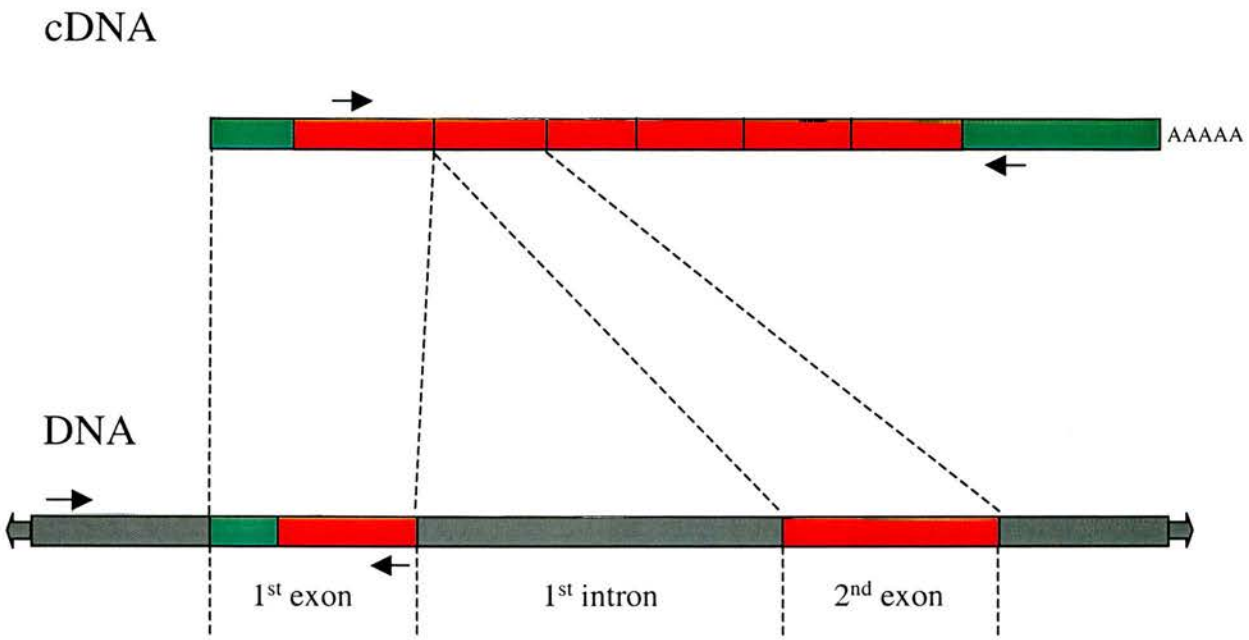
The human sequence differs from the domestic pig at 74 nucleotides and an insertion of nine bases after base 887 (encoding an additional three residues). Fifteen of these differences encode alternative amino acids (Figure 3.5b).

3.4 NFAT2 Regulatory Domain

A 684bp region comprising the regulatory domain of NFAT2 was amplified from cDNA by PCR. This sequence includes the calcineurin-binding region mimicked by A238L (Figure 3.6). This amplicon encodes a primary protein 228 amino acids in length. Eighteen nucleotide differences were identified between the suid species considered here (Table 3.5a). All of these differences are silent between the domestic pig and warthog. Again, however, the babirusa is different, with two mutations at the

Figure 3.4 – IκBα PCR Strategy and Primer Sites

The majority of the IκBα coding sequence was achieved in the normal manner from cDNA. However, due to the presence of a GC-rich 5' untranslated region, the forward primer was designed within the 5' coding sequence of the first exon. To achieve full coding sequence, the remaining 5' region was amplified directly from DNA using a forward primer designed in the promoter region and reverse primer within the first exon, downstream of the forward primer used in the initial PCR from cDNA.



Intronic/promoter DNA

5' or 3' untranslated region

Coding sequence

Forward primer location

Reverse primer location

Figure 3.5 – *IκBα* Comparative Sequence Analysis

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
BB	*Babyrousa babyrussa*	Babirusa
HS	*Homo sapiens*	Human (AY033600)

A *IκBα* complete nucleotide coding sequence. Suid *IκBα* coding sequence is 945 bases long (including the stop codon). Human *IκBα* is 954 base pairs in length, due to nine base pair insertion. Eleven polymorphisms exist between the suids.

B The predicted primary protein sequence for IκBα is 315 amino acids in length in suids and 318 amino acids in humans. The protein sequence is conserved between the domestic pig and warthog, which differ from the babirusa at one residue.

A

```
SS ATG TTC CAG CCC GCA GAG CCC GGC CAG GAG TGG GCC ATG GAG GGG CCC CGG GAC GCG CTC  [ 60]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 60]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 60]
HS ... ... ... G.G ..C ... .G. CC. ... ... ... ... ... ... ..C ... ..C ... .G. ..G  [ 60]

SS AAG AAG GAG CGG CTA CTG GAC GAC CGC CAC GAC AGC GGC CTG GAC TCC ATG AAG GAC GAG  [120]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [120]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [120]
HS ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ...  [120]

SS GAG TAC GAG CAG ATG GTG AAG GAG CTG CGC GAG ATC CGC CTC GAG CCG CAG GAG GCG CCC  [180]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [180]
BB ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ...  [180]
HS ... ... ... ... ... ..C ... ... ... .AG ... ... ... ... ... ... ... ... .T. ..G  [180]

SS CGC GGC GCC GAG CCC TGG AAG CAG CAG CTC ACC GAG GAC GGA GAC TCG TTC CTG CAC TTG  [240]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ...  [240]
BB ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ...  [240]
HS ... ... T.G ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ...  [240]

SS GCC ATC ATC CAT GAA GAG AAG GCA CTG ACC ATG GAA GTG GTC CGC CAA GTG AAG GGA GAT  [300]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [300]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [300]
HS ... ... ... ... ... ..A ... ... ... ... ... ... ... A.. ... ..G ... ... ... ..C  [300]

SS CTG GCT TTT CTT AAC TTC CAG AAC AAC CTG CAG CAG ACT CCA CTC CAC TTG GCG GTG ATC  [360]
PA ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [360]
BB ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [360]
HS ... ..C ..C ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ...  [360]

SS ACC AAC CAG CCA GAA ATC GCT GAG GCA CTT CTG GAA GCT GGC TGT GAT CCT GAG CTC CGA  [420]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [420]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [420]
HS ... ... ... ... ... ..T ... ... ... ... ... .G. ... ... ... ... ... ... ... ...  [420]

SS GAC TTT CGA GGA AAT ACC CCT CTA CAC CTT GCC TGT GAG CAG GGC TGC CTG GCC AGT GTG  [480]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [480]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [480]
HS ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ..C ...  [480]

SS GGA GTC CTG ACT CAG CCC CGC GGG ACC CAG CAC CTC CAC TCC ATT CTG CAG GCC ACC AAC  [540]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [540]
BB ... ... ... ... ... ... ... ... ... ... G.. ... ... ... ... ..C ... ... ... ...  [540]
HS ... ... ... ... ... T.. T.. ACC ... .C. ... ... ... ... ..C ... A.. ..T ... ...  [540]

SS TAC AAT GGC CAC ACA TGT CTG CAC TTA GCC TCG ATC CAT GGC TAC CTG GGC ATT GTG GAG  [600]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [600]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ...  [600]
HS ... ... ... ... ..G ..C ..A ... ... ... ..T ... ... ... ... ... ... ..C ... ...  [600]

SS CTG TTG GTG TCT TTG GGT GCT GAT GTC AAC GCT CAG GAG CCC TGC AAT GGC CGA ACC GCC  [660]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [660]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T  [660]
HS ..T ... ... ..C ... ... ... ... ... ..T ... ... ... ... ..T ... ... ..G ..T ...  [660]

SS CTG CAT CTT GCG GTG GAC CTG CAG AAT CCC GAC CTG GTG TCG CTC TTG TTG AAG TGT GGG  [720]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ...  [720]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [720]
HS ..T ..C ..C ..A ... ... ... ..A ... ..T ... ... ... ..A ... C.. ... ... ... ...  [720]

SS GCT GAT GTC AAC AGA GTC ACC TAC CAG GGC TAC TCC CCG TAC CAG CTC ACC TGG GGC CGC  [780]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [780]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [780]
HS ... ... ... ... ... ..T ... ... ... ... ..T ..T ..C ... ... ... ... ... ... ...  [780]

SS CCA AGC ACT CGG ATA CAG CAG CAG CTG GGC CAG CTG ACC CTA GAA AAC CTC CAG ATG CTT  [840]
PA ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [840]
BB ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [840]
HS ... ... ..C ... ... ... ... ... ... ... ... ... ..A ... ... ... ..T ... ... ..G  [840]
```

```
SS CCA GAG AGC GAG GAT GAG GAG AGC TAT GAC ACG GAG TCA GAG TTC AC- --- --- --A GAG  [900]
PA ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ..- --- --- --. ...  [900]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..- --- --- --. ...  [900]
HS ... ... ..T ... ... ... ... ... ... ... ..A ... ... ... ... ..G GAG TTC AC. ...  [900]

SS GAT GAG CTG CCC TAT GAC GAC TGC GTG CTT GGA GGC CAG CGC CTG ACG TTA TGA  [954]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [954]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [954]
HS ..C ... ... ... ... ..T ... ..T ... T.. ... ... ... ..T ... ... ... ...  [954]
```

B

```
SS MFQPAEPGQE WAMEGPRDAL KKERLLDDRH DSGLDSMKDE EYEQMVKELR EIRLEPQEAP RGAEPWKQQL  [ 70]
PA .......... .......... .......... .......... .......... .......... ..........  [ 70]
BB .......... .......... .......... .......... .......... .......... ..........  [ 70]
HS ...A..RP.. ........G. .......... .......... .........Q ........V. ..S.......  [ 70]

SS TEDGDSFLHL AIIHEEKALT MEVVRQVKGD LAFLNFQNNL QQTPLHLAVI TNQPEIAEAL LEAGCDPELR  [140]
PA .......... .......... .......... .......... .......... .......... ..........  [140]
BB .......... .......... .......... .......... .......... .......... ..........  [140]
HS .......... .......... ...I...... .......... .......... .......... .G........  [140]

SS DFRGNTPLHL ACEQGCLASV GVLTQPRGTQ HLHSILQATN YNGHTCLHLA SIHGYLGIVE LLVSLGADVN  [210]
PA .......... .......... .......... .......... .......... .......... ..........  [210]
BB .......... .......... .........E .......... .......... .......... ..........  [210]
HS .......... .......... .....SCT.P ......K... .......... .......... ..........  [210]

SS AQEPCNGRTA LHLAVDLQNP DLVSLLLKCG ADVNRVTYQG YSPYQLTWGR PSTRIQQQLG QLTLENLQML  [280]
PA .......... .......... .......... .......... .......... .......... ..........  [280]
BB .......... .......... .......... .......... .......... .......... ..........  [280]
HS .......... .......... .......... .......... .......... .......... ..........  [280]

SS PESEDEESYD TESEF---TE DELPYDDCVL GGQRLTL* [318]
PA .......... .....---.. .......... ........ [318]
BB .......... .....---.. .......... ........ [318]
HS .......... .....TEF.. .........F ........ [318]
```

Figure 3.6 – *NFAT2 Regulatory Domain* Comparative Sequence Analysis

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
BB	*Babyrousa babyrussa*	Babirusa
HS	*Homo sapiens*	Human (U08015)

A The nucleotide sequence for the *NFAT2* regulatory domain. A 648bp region encompassing the *NFAT2* regulatory domain was sequenced; this includes 18 polymorphisms between the suids.

B The NFAT2 amplicon encoded a predicted primary protein 228 amino acids in length. The protein sequence is conserved between the domestic pig and warthog, which differ from the babirusa at two residues. However, the consensus CnA-binding domain is entirely conserved between species (double-underlined SPRIEIT).

A

```
SS GAC GGC GGG CCC TCG GGC TAT TTC CTC CCC TCC GGT GGT GTC AGG CCC AAC GGG GCC CCC  [ 60]
PA ... ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 60]
BB ... ... ... ... ..A ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 60]
HS ... ..T ... ... G.. ... ..C ... ... T.. ... ..C CAC AC. ... ..T G.T ... ... ..T  [ 60]

SS GCC CTG GAG AGC CCT CGC ATT GAG ATC ACA TCC TAC CTG GGG CTG CAC CAC AAC AAC GGC  [120]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [120]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [120]
HS ... ... ... ..T ... ... ..C ... ..A ..C ..G .G. T.. ..C ... T.. ... ... ..T AA.  [120]

SS CAG TTC TTC CAT GAT GTG GCT GTG GAA GAC GTT CTT CCC AAC CCC AGA CGC TCC CCG TCC  [180]
PA ... ... ... ... ..C ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ...  [180]
BB ... ... ... ... ..C ... ..C ... ... ..T ... ... ... ... ... ... ... ... ... ...  [180]
HS ... ..T ... ..C ... ... .AG ... ... ... ..C ..C ..T .G. T.. .A. ..G ... ..C ...  [180]

SS ACG GCC ACC TTG AGT CTT CCC AAC CTG GAG GCC TAC CGA GAC CCC TCA TGC CTG AGC CCA  [240]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G  [240]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ..G  [240]
HS ... ... ..G C.. ... ..G ... .G. ... ... ... ... A.. ... ... ..G ... ... ... ..G  [240]

SS GCC AGC AGC CTG TCG TCC CGT AGC TGC AAC TCC GAG GCC TCG TCC TAC GAG TCC AGC TTC  [300]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [300]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [300]
HS ... ... ... ... ..C ... ..G ... ... ... ..A ... ... ..C ... ... ... ... .A. .A.  [300]

SS TCA TAC CCG TAC GCG TCC CCG CAG ACA TCC CCG TGG CAG TCT CCC TGC GTG TCC CCC AAG  [360]
PA ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ...  [360]
BB ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ...  [360]
HS ..G ... ... ... ... ... ..C ... ..G ..G ..A ... ... ... ... ... ... ..T ... ...  [360]

SS ACC ACG GAC CCT GAG GAG GGC TTT CCC CGT GGC CTG GGG GCC TGC AGC CTG CTG GGG TCC  [420]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [420]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [420]
HS ... ... ... ..C ... ... ... ... ... ..C ..G ... ... ... ... .CA ... ... ..T ...  [420]

SS CCG CGG CAC TCA CCG TCC ACC TCG CCC CGC ACC AGT GTT ACC GAG GAG AGC TGG TTG GGG  [480]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ...  [480]
BB ..T ... ... ... ... ... ... ... ... ... ... ... ... ..C ... ... ... ... ... ...  [480]
HS ... .A. ... ..C ..C ... ... ... ... ... G.. ..C ..C ..T ... ... ... ... C.. ..T  [480]

SS GCC CGC ACG TCC AGG CCC TCG TCC CCC TGC AAC AAG AGG AAG TAC GGT CTC AAC GGC CGG  [540]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [540]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ...  [540]
HS ... ... T.C ... ..A ... G.. ... ..T ... ... ... ... ... ... A.C ... ... ... ...  [540]

SS CAA CTG TCC TGC TCC CCG CAT GCC TCG CCC ACA CCT TCC CCG CAC AGC TCG CCG AGG GTC  [600]
PA ... ... ... ... ... ..A ..C ... ... ... ..G ..G ... ... ... ... ... ... ... ...  [600]
BB ... ... ... ... ... ... ..C ... ... ... ..G ..G ... ... ... ... ... ... ... ...  [600]
HS ..G .C. C.. .A. ..A ..C ..C CA. ... ... ..G ..G ... ... ... G.. ..C ... C.. ...  [600]

SS AGC GTC ACC GAC GAC ACT TGG CTG GGC AAC ACC ACG CAG TAC ACG AGC TCT GCC ATC GTG  [660]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [660]
BB ... A.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [660]
HS ... ..G ... ... ... T.G ... T.. ... ... ... ..C ... ... ..C ... ..G ... ... ...  [660]

SS GCA GCC ATC AAT GCC CTG AGC ACC  [684]
PA ..T ... ... ... ... ... ... ...  [684]
BB ..T ... ... ... ... ... ... .A.  [684]
HS ..C ... ... ..C ..G ... .C. ...  [684]
```

B

```
SS DGGPSGYFLP SGGVRPNGAP ALESPRIEIT SYLGLHHNNG QFFHDVAVED VLPNPRRSPS TATLSLPNLE  [ 70]
PA .......... .......... .......... .......... .......... .......... ..........  [ 70]
BB .......... .......... .......... .......... .......... .......... ..........  [ 70]
HS ....A....S ..HT..D... .......... .C...Y...N ......E... ...SSK.... .......S..  [ 70]

SS AYRDPSCLSP ASSLSSRSCN SEASSYESSF SYPYASPQTS PWQSPCVSPK TTDPEEGFPR GLGACSLLGS  [140]
PA .......... .......... .......... .......... .......... .......... ..........  [140]
BB .......... .......... .......... .......... .......... .......... ..........  [140]
HS .......... .......... ........NY .......... .......... .......... .....T....  [140]

SS PRHSPSTSPR TSVTEESWLG ARTSRPSSPC NKRKYGLNGR QLSCSPHASP TPSPHSSPRV SVTDDTWLGN  [210]
PA .......... .......... .......... .......... .......... .......... ..........  [210]
BB .......... .......... .......... .......... .......... .......... .I........  [210]
HS .Q........ A......... ..S...A... .....S.... .PPY...H.. .....G.... .....S....  [210]

SS TTQYTSSAIV AAINALST [228]
PA .......... ........ [228]
BB .......... .......N [228]
HS .......... ......T. [228]
```

Table 3.4

A The eleven nucleotide differences existing between domestic pig, warthog and babirusa *IκBα* sequence.

B The one amino acid difference between the babirusa IκBα primary protein sequence and the other suids. Primary protein sequence is conserved between the domestic pig and warthog.

A

Nucleotide Position	Domestic pig	Warthog	Babirusa	Codon Position	Silent mutation? Which species differs?
150	C	C	G	3rd	Silent
213	C	C	G	3rd	Silent
219	C	T	C	2nd	Silent
312	T	C	C	3rd	Silent
508	C	C	G	1st	Babirusa
525	T	T	C	3rd	Silent
594	T	T	C	3rd	Silent
657	T	T	C	3rd	Silent
702	G	A	G	3rd	Silent
789	T	C	C	3rd	Silent
858	G	A	G	3rd	Silent

B

Amino Acid Position	Corresponding codon	Domestic pig	Warthog	Babirusa
170	508-510	Glutamine (Q)	Glutamine (Q)	Glutamic acid (E)

Table 3.5

A The 18 nucleotide differences existing between domestic pig, warthog and babirusa *NFAT2* regulatory domain sequence.

B The two amino acid differences between the babirusa NFAT2 primary protein sequence and the other suids. Primary protein sequence is conserved between the domestic pig and warthog.

A

Nucleotide Position	Domestic pig	Warthog	Babirusa	Codon Position	Silent mutation? Which species differs?
15	G	G	A	3rd	Silent
21	T	C	C	3rd	Silent
135	T	C	C	3rd	Silent
141	T	C	C	3rd	Silent
150	C	C	T	3rd	Silent
228	A	A	G	3rd	Silent
240	A	G	G	3rd	Silent
327	A	G	G	3rd	Silent
423	G	G	T	3rd	Silent
459	T	C	C	3rd	Silent
519	G	G	A	3rd	Silent
558	G	A	G	3rd	Silent
561	T	C	C	3rd	Silent
573	A	G	G	3rd	Silent
576	T	G	G	3rd	Silent
604	G	G	A	1st	Babirusa
663	A	T	T	3rd	Silent
683	C	C	A	2nd	Babirusa

B

Amino Acid Position	Corresponding codon	Domestic pig	Warthog	Babirusa
202	604-606	Valine (V)	Valine (V)	Isoleucine (I)
228	682-684	Threonine (T)	Threonine (T)	Asparagine (N)

amino acid level - an isoleucine (as opposed to valine) at position 202 and an asparagine (as opposed to threonine) at position 228 (Table 3.5b).

The human sequence differs from the domestic pig at 104 nucleotides with 28 of these differences encoding alternative amino acids. However, the calcineurin-binding sequence remains conserved between all suid and human sequences (Figure 3.6b).

3.5 Calcineurin Aβ

Calcineurin Aβ (CnAβ) mRNA sequence is 1545bp in length in the domestic pig and babirusa and encodes a primary protein 515 amino acids long. However, CnAβ mRNA in the warthog is 1578bp long and encodes a primary protein 526 amino acids in length (Figure 3.7). The warthog CnAβ nucleotide sequence differs from the domestic pig and babirusa in having a 3bp and a 30bp insertion - the 3bp insertion in warthog adds a single valine residue (V), and the 30bp insertion adds ten residues (ATVEAIEAEK). In humans, two CnAβ splice variants exist. Splice variant 1 includes a 3bp insertion and splice variant 2 encodes a 30bp insertion. However, these two insertion events appear to occur simultaneously within a single warthog transcript.

Despite numerous attempts, the 5' sequence of CnAβ, corresponding to the first 118bp of human sequence, could not be amplified in any of the suid species. This 5' encodes a GC-rich polyproline domain through which it appears the PCR reaction failed to extend. Outwith the 5' region, and in addition to warthog insertions, seven remaining nucleotide differences were identified between the suids (Table 3.6). All seven differences are silent.

The two warthog insertions occur outside the known functional domains of CnAβ. The single valine insertion occurs between the calcineurin B-binding domain and the calmodulin-binding domain and the ten amino acid insertion occurs between the calmodulin-binding and autoinhibitory domains (Figure 3.7b).

Figure 3.7 – *Calcineurin Aβ* Comparative Sequence Analysis

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
BB	*Babyrousa babyrussa*	Babirusa
HS1	*Homo sapiens* 'PPP3CB'	Human (M29551)
HS2	*Homo sapiens* 'Q8N3W4'	Human (AJ488506)

A *Calcineurin Aβ* (*CnAβ*) nucleotide coding sequence. *CnAβ* nucleotide sequence is 1545 bases long in the domestic pig and babirusa and 1578 in the warthog. The first 118bp are absent from the pig sequences. The warthog contains two insertions, one is three base pairs in length and the other is thirty base pairs in length. Human transcript PPP3CB contains an equivalent thirty base pair insertion, whereas transcript Q8N3W4 has the three base pair insertion.

B Calcineurin Aβ predicted primary protein structure without the first 40 amino acids of pig sequence, which encodes a polyproline N-terminal domain (as seen in the human sequences). CnAβ protein sequence is conserved between all pigs, except for the two insertions in the warthog sequence: a single valine insertion and a ten amino acid insertion (ATVEAIEAEK). These insertions occur between the four functional domains: Catalytic domain single-underline; calcineurin B-binding domain dashed-underline; calmodulin-binding domain double-underline; autoinhibitory domain dotted-underline.

A

```
SS  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [  60]
PA  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [  60]
BB  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- ---  [  60]
HS1 ATG GCC GCC CCG GAG CCG GCC CGG GCT GCA CCG CCC CCA CCC CCG CCC CCG CCG CCC CCT  [  60]
HS2 ATG GCC GCC CCG GAG CCG GCC CGG GCT GCA CCG CCC CCA CCC CCG CCC CCG CCG CCC CCT  [  60]

SS  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -CT  [ 120]
PA  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -.C  [ 120]
BB  --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- --- -.C  [ 120]
HS1 CCC GGG GCT GAC CGC GTC GTC AAA GCT GTC CCT TTC CCC CCA ACA CAT CGC TTG ACA T..  [ 120]
HS2 CCC GGG GCT GAC CGC GTC GTC AAA GCT GTC CCT TTC CCC CCA ACA CAT CGC TTG ACA T..  [ 120]

SS  GAA GAA GTA TTT GAC ATG GAT GGG ATA CCC AGG GTT GAT GTT CTG AAG AAC CAC TTA GTA  [ 180]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 180]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 180]
HS1 ... ... ... ... ..T T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ..G  [ 180]
HS2 ... ... ... ... ..T T.. ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ..G  [ 180]

SS  AAA GAA GGT CGG GTG GAT GAA GAA ATT GCA CTT AGA ATT ATC AAT GAG GGT GCT GCC ATT  [ 240]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 240]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 240]
HS1 ... ... ... ..A ..A ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ..C  [ 240]
HS2 ... ... ... ..A ..A ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ..C  [ 240]

SS  CTT CGG AGA GAG AAA ACC ATG ATA GAA GTA GAG GCT CCA ATT ACA GTG TGT GGT GAC ATC  [ 300]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 300]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 300]
HS1 ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ..C ... ... ... ... ... ... ...  [ 300]
HS2 ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ..C ... ... ... ... ... ... ...  [ 300]

SS  CAT GGC CAA TTT TTT GAT CTG ATG AAA CTT TTT GAA GTA GGA GGA TCA CCT GCT AAT ACA  [ 360]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 360]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 360]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 360]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 360]

SS  CGA TAC CTT TTT CTT GGT GAT TAT GTG GAC AGA GGT TAT TTT AGT ATA GAG TGT GTC TTA  [ 420]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 420]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 420]
HS1 ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 420]
HS2 ... ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 420]

SS  TAT TTA TGG GTC CTG AAG ATT CTA TAC CCA AGC ACA TTA TTT CTT CTG AGA GGC AAC CAT  [ 480]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 480]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 480]
HS1 ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 480]
HS2 ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 480]

SS  GAA TGC AGA CAC CTT ACT GAA TAT TTT ACC TTT AAA CAG GAA TGT AAA ATT AAA TAT TCA  [ 540]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 540]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 540]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ..G ... ..G  [ 540]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ..G ... ..G  [ 540]

SS  GAA AGA GTC TAT GAA GCT TGT ATG GAA GCT TTT GAT AGC CTC CCT CTT GCT GCA CTT TTA  [ 600]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 600]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 600]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ..T T.G ... ... ... ... ... ...  [ 600]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ..T T.G ... ... ... ... ... ...  [ 600]

SS  AAT CAA CAA TTT CTT TGT GTT CAT GGT GGA CTT TCA CCA GAA ATA CAC ACA CTG GAT GAT  [ 660]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 660]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 660]
HS1 ..C ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 660]
HS2 ..C ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 660]

SS  ATT AGA AGA TTA GAT AGA TTC AAA GAG CCA CCT GCA TTT GGA CCA ATG TGT GAC TTG CTA  [ 720]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 720]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 720]
HS1 ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... T..  [ 720]
HS2 ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... T..  [ 720]
```

```
SS  TGG TCT GAT CCT TCT GAA GAT TTT GGA AAT GAA AAA TCA CAG GAA CAT TTT AGT CAC AAT  [ 780]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ...  [ 780]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ... ...  [ 780]
HS1 ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 780]
HS2 ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 780]

SS  ACA GTT CGA GGA TGT TCT TAT TTT TAT AAC TAT CCA GCA GTG TGT GAA TTT TTG CAA AAC  [ 840]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 840]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 840]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 840]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 840]

SS  AAT AAT TTG TTA TCG ATT ATT AGA GCT CAT GAA GCT CAA GAT GCA GGC TAT AGA ATG TAC  [ 900]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 900]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 900]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 900]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 900]

SS  AGA AAA AGT CAA ACT ACA GGG TTC CCT TCA TTA ATA ACA ATT TTT TCG GCA CCT AAT TAC  [ 960]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 960]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 960]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 960]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 960]

SS  TTA GAT GTC TAC AAT AAT AAA GCT GCT GTA TTA AAG TAT GAA AAT AAT GTG ATG AAT ATT  [1020]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1020]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1020]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1020]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1020]

SS  CGA CAA TTT AAC TGT TCT CCA CAT CCT TAC TGG TTG CCC AAT TTT ATG GAT GTC TTC ACC  [1080]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1080]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1080]
HS1 ... ..G ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ..G  [1080]
HS2 ... ..G ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ..G  [1080]

SS  TGG TCC TTA CCA TTT GTT GGA GAA AAA GTG ACA GAA ATG TTG GTA AAT GTT CTC AGT ATT  [1140]
PA  ... ..T ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ..C ... ... ...  [1140]
BB  ... ..T ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ...  [1140]
HS1 ... ..T ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ...  [1140]
HS2 ... ..T ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ...  [1140]

SS  TGC TCT GAT GAT GAA CTG ATG ACT GAA GGT GAA GAC CAG TTT GAT G-- -GT TCA GCT GCA  [1200]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .TA G.. ... ... ...  [1200]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .-- -.. ... ... ...  [1200]
HS1 ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... .-- -.. ... ... ...  [1200]
HS2 ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... .TA G.. ... ... ...  [1200]

SS  GCC CGG AAA GAA ATT ATA AGA AAC AAA ATT CGA GCA ATT GGC AAG ATG GCA AGA GTC TTC  [1260]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ...  [1260]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ... ...  [1260]
HS1 ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1260]
HS2 ... ... ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1260]

SS  TCT GTT CTC AGG GAG GAG AGT GAA AGT GTG CTG ACA CTC AAG GGC CTG ACT CCC ACA GGG  [1320]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1320]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1320]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1320]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1320]

SS  ATG TTA CCT AGT GGA GTG TTG GCT GGA GGA CGG CAG ACC CTG CAA AGT GC- --- --- ---  [1380]
PA  ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ACA GTT GAG  [1380]
BB  ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..- --- --- ---  [1380]
HS1 ... ..G ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ..C ACA GTT GAG  [1380]
HS2 ... ..G ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ..- --- --- ---  [1380]

SS  --- --- --- --- --- --- --A ATA CGA GGA TTC TCT CCA CCA CAT AGA ATC TGC AGT TTT  [1440]
PA  GCT ATT GAG GCT GAA AAA GC. ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1440]
BB  --- --- --- --- --- --- --. ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1440]
HS1 GCT ATT GAG GCT GAA AAA GC. ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1440]
HS2 --- --- --- --- --- --- --. ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1440]

SS  GAA GAG GCA AAG GGT TTG GAT AGG ATC AAT GAG AGA ATG CCA CCC CGG AAA GAT GCT GTT  [1500]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1500]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1500]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ..A  [1500]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ..A  [1500]
```

```
SS  CAG CAA GAT GGT TTC AAT TCT CTG AAC ACC GCA CAT GCC ACT GAG AAC CAC GGG ACT GGC  [1560]
PA  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1560]
BB  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1560]
HS1 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ...  [1560]
HS2 ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G ...  [1560]

SS  AAC CAT GGT GCC CAG TGA  [1578]
PA  ... ... ... ... ... ...  [1578]
BB  ... ... ... ... ... ...  [1578]
HS1 ... ... AC. ... ... ...  [1578]
HS2 ... ... AC. ... ... ...  [1578]
```

B

```
SS  ---------- ---------- ---------- ---------- EEVFDMDGIP RVDVLKNHLV KEGRVDEEIA  [ 70]
PA  ---------- ---------- ---------- ---------- .......... .......... ..........  [ 70]
BB  ---------- ---------- ---------- ---------- .......... .......... ..........  [ 70]
HS1 MAAPEPARAA PPPPPPPPPP PGADRVVKAV PFPPTHRLTS .....L.... .......... ..........  [ 70]
HS2 MAAPEPARAA PPPPPPPPPP PGADRVVKAV PFPPTHRLTS .....L.... .......... ..........  [ 70]

SS  LRIINEGAAI LRREKTMIEV EAPITVCGDI HGQFFDLMKL FEVGGSPANT RYLFLGDYVD RGYFSIECVL  [140]
PA  .......... .......... .......... .......... .......... .......... ..........  [140]
BB  .......... .......... .......... .......... .......... .......... ..........  [140]
HS1 .......... .......... .......... .......... .......... .......... ..........  [140]
HS2 .......... .......... .......... .......... .......... .......... ..........  [140]

SS  YLWVLKILYP STLFLLRGNH ECRHLTEYFT FKQECKIKYS ERVYEACMEA FDSLPLAALL NQQFLCVHGG  [210]
PA  .......... .......... .......... .......... .......... .......... ..........  [210]
BB  .......... .......... .......... .......... .......... .......... ..........  [210]
HS1 .......... .......... .......... .......... .......... .......... ..........  [210]
HS2 .......... .......... .......... .......... .......... .......... ..........  [210]

SS  LSPEIHTLDD IRRLDRFKEP PAFGPMCDLL WSDPSEDFGN EKSQEHFSHN TVRGCSYFYN YPAVCEFLQN  [280]
PA  .......... .......... .......... .......... .......... .......... ..........  [280]
BB  .......... .......... .......... .......... .......... .......... ..........  [280]
HS1 .......... .......... .......... .......... .......... .......... ..........  [280]
HS2 .......... .......... .......... .......... .......... .......... ..........  [280]

SS  NNLLSIIRAH EAQDAGYRMY RKSQTTGFPS LITIFSAPNY LDVYNNKAAV LKYENNVMNI RQFNCSPHPY  [350]
PA  .......... .......... .......... .......... .......... .......... ..........  [350]
BB  .......... .......... .......... .......... .......... .......... ..........  [350]
HS1 .......... .......... .......... .......... .......... .......... ..........  [350]
HS2 .......... .......... .......... .......... .......... .......... ..........  [350]

SS  WLPNFMDVFT WSLPFVGEKV TEMLVNVLSI CSDDELMTEG EDQFD-GSAA ARKEIIRNKI RAIGKMARVF  [420]
PA  .......... .......... .......... .......... .....V.... .......... ..........  [420]
BB  .......... .......... .......... .......... .....-.... .......... ..........  [420]
HS1 .......... .......... .......... .......... .....-.... .......... ..........  [420]
HS2 .......... .......... .......... .......... .....V.... .......... ..........  [420]

SS  SVLREESESV LTLKGLTPTG MLPSGVLAGG RQTLQS---- ------AIRG FSPPHRICSF EEAKGLDRIN  [490]
PA  .......... .......... .......... ......ATVE AIEAEK.... .......... ..........  [490]
BB  .......... .......... .......... ......---- ------.... .......... ..........  [490]
HS1 .......... .......... .......... ......ATVE AIEAEK.... .......... ..........  [490]
HS2 .......... .......... .......... ......---- ------.... .......... ..........  [490]

SS  ERMPPRKDAV QQDGFNSLNT AHATENHGTG NHGAQ* [526]
PA  .......... .......... .......... ..... [526]
BB  .......... .......... .......... ..... [526]
HS1 .......... .......... .......... ..T.. [526]
HS2 .......... .......... .......... ..T.. [526]
```

Table 3.6

The seven nucleotide differences existing between domestic pig, warthog and babirusa *calcineurin A*β sequence.

Nucleotide Position	Domestic pig	Warthog	Babirusa	Codon Position	Silent mutation? Which species differs?
120	T	C	C	2nd	Silent
765	A	G	G	3rd	Silent
1086	C	T	T	3rd	Silent
1113	A	G	G	3rd	Silent
1131	T	C	T	3rd	Silent
1254	A	G	G	3rd	Silent
1326	A	G	G	3rd	Silent

3.6 p65

p65 coding sequence is 1662bp in length in the domestic pig and warthog and encodes a primary protein of 554 amino acids. In the babirusa and human, p65 coding sequence is 1656bp in length and encodes a protein of 552 amino acids (Figure 3.8). Full coding sequence was obtained in all three suid species, except for the first six bases, which could not be amplified due to insufficient PCR priming sites within the 5' untranslated region (UTR). Although the babirusa and human sequence are both 6bp shorter than the domestic pig and warthog sequences, these deletions occur at different locations. The babirusa has a single 6bp deletion in the region between the rel homology domain and transactivation 2 domain (TA2), whereas the human has two separate 3bp deletions. One is located 20bp N-terminal to the deletion in the babirusa, the other is located in the TA2, 50bp C-terminal to the deletion in the babirusa (Figure 3.8b). Significantly, between the suid species, 33 nucleotide differences were identified (Table 3.7a) - nine of these differences encode mutations at amino acid level (Table 3.7b). Six of these are silent between the domestic pig and warthog, but differ in the babirusa.

Of particular interest are the three amino acids which differ between p65 sequence in the domestic pig and warthog:

1. A threonine (T) residue at position 448 in the domestic pig, occurs as alanine (A) in warthog and babirusa, but is completely absent in the human (the TA2 deletion site).
2. A serine (S) residue at position 485 in the domestic pig exists as a proline (P) in the other suids and human sequence.
3. A proline (P) residue at position 531 in warthog is conserved as a serine (S) residue in the other pigs and human.

All three of these amino acid mutations occur outside the rel homology domain (RHD), with positions 448 and 485 being located within the transactivation domain 2 (TA2) and position 531 within the transactivation domain 1 (TA1).

Figure 3.8 – *p65* Comparative Sequence Analysis

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
BB	*Babyrousa babyrussa*	Babirusa
HS	*Homo sapiens*	Human (M62399)

A *p65* nucleotide coding sequence, minus the first 6bp in pigs. Suid *p65* coding sequence is 1662 bases long in the domestic pig and warthog and 1656 bases in the babirusa. Nucleotide coding sequence contains 33 polymorphisms between the suids.

B p65 predicted primary protein structure, minus the first 2 amino acids in pigs. The rel homology domain is underlined, the transactivation 2 domain is dashed-underlined and the transactivation 1 domain is double-underlined. The nuclear localisation signal (KRKR) dotted-underlined.

A

```
SS --- --- GAC CTC TTC CCC CTC ATC TTC CCC TCG GAG CCG GCC CCG GCC TCG GGC CCC TAT  [  60]
PA --- --- ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ...  [  60]
BB --- --- ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [  60]
HS ATG GAC ..A ..G ... ... ... ... ... ..G G.A ... ..A ... .A. ... ..T ... ... ...  [  60]

SS GTG GAG ATC ATC GAG CAG CCC AAG CAG CGG GGC ATG CGC TTC CGC TAC AAG TGC GAG GGC  [ 120]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ...  [ 120]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 120]
HS ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..G  [ 120]

SS CGC TCA GCC GGC AGT ATC CCG GGC GAG AGG AGC ACG GAT ACC ACC AAG ACC CAC CCC ACC  [ 180]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 180]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 180]
HS ... ..C ..G ... ..C ... ..A ... ... ... ... ..A ... ... ... ... ... ... ... ...  [ 180]

SS ATC AAG ATC AAT GGC TAC ACG GGG CCA GGG ACA GTG CGC ATC TCC CTG GTC ACC AAG GAC  [ 240]
PA ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ... ... ... ..T ... ...  [ 240]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ...  [ 240]
HS ... ... ... ... ... ... ..A ..A ... ... ... ... ... ... ... ... ... ... ... ...  [ 240]

SS CCC CCT CAC CGG CCT CAC CCC CAT GAG CTC GTG GGG AAA GAC TGC CGG GAT GGC TTC TAT  [ 300]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 300]
BB ... ... ... ... ... ... ..T ... ... ..T ... ... ... ... ... ... ... ... ... ...  [ 300]
HS ..T ... ... ... ... ... ... ..C ... ..T ..A ..A ..G ... ... ... ... ... ... ...  [ 300]

SS GAG GCT GAG CTC TGC CCA GAC CGC TGC ATC CAC AGC TTC CAG AAC CTG GGG ATC CAG TGT  [ 360]
PA ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 360]
BB ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 360]
HS ... ... ... ... ... ..G ... ... ... ... ... ..T ... ... ... ... ..A ... ... ...  [ 360]

SS GTA AAG AAG CGG GAC CTG GAA CAG GCC ATC AAT CAG CGC ATC CAG ACC AAC AAC AAC CCC  [ 420]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ...  [ 420]
BB ... ... ... ... ... ... ... ... ... ... ... .G. ... ... ..T ... ... ... ..T ... ...  [ 420]
HS ..G ... ... ... ... ... ..G ... ..T ... .G. ... ... ... ... ... ... ... ... ...  [ 420]

SS TTC CAA GTT CCC ATA GAA GAG CAG CGC GGG GAC TAC GAC CTG AAT GCT GTG CGG CTC TGC  [ 480]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 480]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 480]
HS ... ... ... ..T ... ... ... ... ..T ... ... ... ... ... ... ... ... ... ... ...  [ 480]

SS TTC CAG GTG ACA GTG CGG GAC CCA GCA GGC AGG CCC CTC CGC CTG CCG CCT GTC CTC TCT  [ 540]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 540]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ...  [ 540]
HS ... ... ... ... ... ... ... ... T.. ... ... ... ... ... ... ... ... ... ..T C..  [ 540]

SS CAC CCC ATC TTT GAC AAC CGT GCC CCC AAC ACT GCA GAG CTC AAG ATC TGC CGG GTG AAT  [ 600]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ...  [ 600]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 600]
HS ..T ... ... ... ... ..T ... ... ... ... ... ..C ... ... ... ... ... ..A ... ..C  [ 600]

SS CGG AAC TCG GGG AGC TGC CTT GGG GGC GAT GAG ATC TTC CTG CTG TGC GAC AAG GTG CAG  [ 660]
PA ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 660]
BB ... ... ..C ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 660]
HS ..A ... ..T ..C ... ... ..C ..T ..G ... ... ... ... ..A ... ..T ... ... ... ...  [ 660]

SS AAA GAG GAC ATC GAG GTG TAT TTC ACG GGC CCG GGC TGG GAG GCC CGA GGC TCC TTT TCA  [ 720]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 720]
BB G.. ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ... ... ... ... ...  [ 720]
HS ... ... ... ..T ... ... ... ... ... ..A ..A ... ... ... ... ... ... ... ... ..G  [ 720]

SS CAA GCC GAC GTG CAC CGA CAA GTG GCC ATC GTG TTC CGG ACG CCT CCC TAC GCG GAC CCC  [ 780]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 780]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... T.. ... ...  [ 780]
HS ... ..T ..T ... ... ... ... ... ... ..T ... ... ... ..C ... ... ... ..A ... ...  [ 780]

SS AGC CTG CAG GCC CCC GTG CGC GTC TCC ATG CAG CTG CGG CGG CCT TCG GAT CGG GAG CTC  [ 840]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 840]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 840]
HS ... ... ... ..T ..T ... ..T ... ... ... ... ... ... ... ... ..C ..C ... ... ...  [ 840]

SS AGC GAG CCC ATG GAA TTC CAG TAC TTG CCA GAC ACA GAT GAC CGG CAC CGG ATT GAG GAG  [ 900]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [ 900]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ..T ... ... ... ..A  [ 900]
HS ..T ... ... ... ... ... ... ... C.. ... ..T ... ..C ..T ..T ... ... ... ... ...  [ 900]
```

```
SS AAA CGC AAA AGG ACC TAT GAG ACC TTT AAG AGC ATC ATG AAG AAG AGT CCT TTC AAT GGA  [ 960]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..C ...  [ 960]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ... ...  [ 960]
HS ... ..T ... ... ..A ... ... ... ..C ... ... ... ... ... ... ... ... ... .GC ...  [ 960]

SS CCC ACC GAC CCC CGG CCT GCA ACC CGG CGC ATT GCT GTG CCT TCC CGC AGC TCA GCT TCC  [1020]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1020]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ...  [1020]
HS ... ... ... ... ... ... C.. C.T ..A ... ... ... ... ... ... ... ... ... ... ..T  [1020]

SS GTC CCC AAG CCA GCT CCC CAG CCC TAT CCC TTT ACG CCA TCT CTC AGC ACC ATC AAC TTT  [1080]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1080]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1080]
HS ... ... ... ... ..A ... ... ... ... ... ... ... T.. ..C ..G ... ... ... ... .A.  [1080]

SS GAC GAG TTC ACG CCC ATG GCC TTT GCT TCT GGG CAG ATC CCA GGC CAG ACC TCA GCC TTG  [1140]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1140]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... G.. ... ...  [1140]
HS ..T ... ..T C.C A.. ... .TG ... C.. ... ... ... ... --- A.. ... G.. ..G ... ...  [1140]

SS GCC CCA GCC CCT GCC CCA GTC CTG GTC CAG GCC CCA GCC CCG GCC CCA GCC CCA GCC ATG  [1200]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1200]
BB .T. ... ... ... ... ... ... ... ... ... ... ... ..G ... ... ... ..G ..- --- --. ...  [1200]
HS ... ..G ... ... C.. .A. ... ... CC. ... ..T ... ... ..T ... ..T ..T ... ... ...  [1200]

SS GCA TCA GCT CTG GCC CAG GCC CCA GCC CCT GTC CCC GTC CTA GCC CCC GGC CTT GCT CAG  [1260]
PA ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ... ...  [1260]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1260]
HS .T. ... ... ... ... ... ... ... ... ... ... ... ..A ... ... ... ..A ... .C. C.. ...  [1260]

SS GCT GTG GCC CCG CCT GCC CCT AAA ACC AAC CAG GCT GGG GAA GGG ACA CTG ACA GAG GCC  [1320]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1320]
BB ... ... ... ... ... ... ... ... ..G ... ... ... ... ... ... ... ... ... ... ...  [1320]
HS ... ... ... ..A ... ... ..C ..G C.. .C. ... ... ... ... ..A ..G ... T.. ... ...  [1320]

SS CTG CTG CAG CTG CAG TTT GAT ACT GAT GAG GAC CTG GGG GCC CTG CTC GGC AAT AAC ACT  [1380]
PA ... ... ... ... ... ... ... G.. ... ... ... ... ... ... ... ... ... ... ... ...  [1380]
BB ... ... ... ... ... ... ... G.. ... ... ... ... ... ... ... ... ... ... ... ...  [1380]
HS ... ... ... ... ... ... ... --- ... ..A ... ... ... ... T.. ..T ... ..C .G. ..A  [1380]

SS GAC CCG ACC GTG TTC ACG GAC CTG GCA TCC GTC GAC AAC TCT GAG TTT CAG CAG CTG CTG  [1440]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1440]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... .C. ... ... ...  [1440]
HS ... ..A G.T ... ... ..A ... ... ... ... ... ... ... ..C ... ... ... ... ... ...  [1440]

SS AAC CAG GGT GTA TCC ATG CCC CCC CAC ACA GCT GAG CCC ATG CTG ATG GAG TAC CCT GAG  [1500]
PA ... ... ... ... C.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1500]
BB ... ... ... ... C.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1500]
HS ... ... ..C A.. C.T G.. G.. ... ... ... A.. ... ... ... ... ... ... ... ... ...  [1500]

SS GCT ATA ACT CGC TTG GTG ACA GGG TCC CAG AGA CCC CCT GAC CCA GCT CCC ACT CCC CTG  [1560]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1560]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1560]
HS ... ... ... ... C.A ... ... ... G.. ... ..G ... ..C ... ... ... ..T G.. ..A ...  [1560]

SS GGG GCC TCT GGG CTC ACC AAC GGT CTC CTC TCG GGG GAC GAA GAC TTC TCC TCC ATT GCG  [1620]
PA ... ... ... ... ... ... ... ... ... ... ... C.. ... ... ... ... ... ... ... ...  [1620]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ..T ... ... ... ... ... ...  [1620]
HS ... ... C.G ... ... C.. ..T ..C ... ..T ..A ..A ..T ... ... ... ... ... ... ...  [1620]

SS GAC ATG GAC TTC TCA GCC CTT CTG AGT CAG ATC AGC TCC TAA  [1662]
PA ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1662]
BB ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1662]
HS ... ... ... ... ... ... ..G ... ... ... ... ... ... ...  [1662]
```

B

```
SS --DLFPLIFP SEPAPASGPY VEIIEQPKQR GMRFRYKCEG RSAGSIPGER STDTTKTHPT IKINGYTGPG  [ 70]
PA --......... .......... .......... .......... .......... .......... ..........  [ 70]
BB --......... .......... .......... .......... .......... .......... ..........  [ 70]
HS MDE....... A...Q..... .......... .......... .......... .......... ..........  [ 70]

SS TVRISLVTKD PPHRPHPHEL VGKDCRDGFY EAELCPDRCI HSFQNLGIQC VKKRDLEQAI NQRIQTNNNP  [140]
PA .......... .......... .......... .......... .......... .......... ..........  [140]
BB .......... .......... .......... .......... .......... .......... S.........  [140]
HS .......... .......... .......... .......... .......... .......... S.........  [140]

SS FQVPIEEQRG DYDLNAVRLC FQVTVRDPAG RPLRLPPVLS HPIFDNRAPN TAELKICRVN RNSGSCLGGD  [210]
PA .......... .......... .......... .......... .......... .......... ..........  [210]
BB .......... .......... .......... .......... .......... .......... ..........  [210]
HS .......... .......... .........S. .........P .......... .......... ..........  [210]

SS EIFLLCDKVQ KEDIEVYFTG PGWEARGSFS QADVHRQVAI VFRTPPYADP SLQAPVRVSM QLRRPSDREL  [280]
PA .......... .......... .......... .......... .......... .......... ..........  [280]
BB .......... E......... .......... .......... .......S.. .......... ..........  [280]
HS .......... .......... .......... .......... .......... .......... ..........  [280]

SS SEPMEFQYLP DTDDRHRIEE KRKRTYETFK SIMKKSPFNG PTDPRPATRR IAVPSRSSAS VPKPAPQPYP  [350]
PA .......... .......... .......... .......... .......... .......... ..........  [350]
BB .......... .......... .......... .......... .......... .......... ..........  [350]
HS .......... .......... .......... ........S. ......PP.. .......... ..........  [350]

SS FTPSLSTINF DEFTPMAFAS GQIPGQTSAL APAPAPVLVQ APAPAPAPAM ASALAQAPAP VPVLAPGLAQ  [420]
PA .......... .......... .......... .......... .......... .......... ..........  [420]
BB .......... .......... ........A.. V......... ........--. .......... ..........  [420]
HS ..S......Y ...PT.V.P. ...-S.A... ....PQ..P. .......... V......... .......PP.  [420]

SS AVAPPAPKTN QAGEGTLTEA LLQLQFDTDE DLGALLGNNT DPTVFTDLAS VDNSEFQQLL NQGVSMPPHT  [490]
PA .......... .......... ........A.. .......... .......... .......... ....P.....  [490]
BB .......... .......... ........A.. .......... .......... .....S.... ....P.....  [490]
HS ........PT .......S.. ........-.. ........S. ..A....... .......... ...IPVA...  [490]

SS AEPMLMEYPE AITRLVTGSQ RPPDPAPTPL GASGLTNGLL SGDEDFSSIA DMDFSALLSQ ISS*  [554]
PA .......... .......... .......... .......... P......... .......... ....  [554]
BB .......... .......... .......... .......... .......... .......... ....  [554]
HS T......... ........A. ........A.. ..P..P.... .......... .......... ....  [554]
```

Table 3.7

A The 33 nucleotide differences existing between domestic pig, warthog and babirusa *p65* sequence.

B The nine amino acid differences between the domestic pig, warthog and babirusa p65 primary protein sequence. Of particular interest are the three amino acids which differ between the domestic pig and warthog: The threonine (T) residue at position 448 in the domestic pig occurs as alanine (A) in the warthog and babirusa. A serine (S) residue at position 485 in the domestic pig exists as a proline (P) in the warthog and babirusa. The proline (P) residue at position 531 in warthog p65 is a serine (S) residue in the warthog and babirusa.

A

Nucleotide Position	Domestic pig	Warthog	Babirusa	Codon Position	Silent mutation? Which species differs?
51	C	A	C	3rd	Silent
114	C	T	C	3rd	Silent
101	G	A	G	3rd	Silent
234	C	T	T	3rd	Silent
261	C	C	T	3rd	Silent
270	C	C	T	3rd	Silent
318	A	G	G	3rd	Babirusa
392	A	A	G	2nd	Silent
402	C	C	T	3rd	Silent
414	C	T	T	3rd	Silent
525	C	C	A	3rd	Silent
591	C	T	C	3rd	Silent
609	G	C	C	3rd	Silent
661	A	A	G	1st	Babirusa
693	G	G	A	3rd	Silent
772	G	G	T	3rd	Babirusa
885	G	G	A	3rd	Silent
887	C	C	T	3rd	Silent
900	G	G	A	3rd	Silent
942	G	G	A	3rd	Silent
957	T	C	T	3rd	Silent
1005	C	C	T	3rd	Silent
1132	T	T	G	1st	Babirusa
1142	C	C	T	2nd	Babirusa
1176	A	A	G	3rd	Silent
1188	A	A	G	3rd	Silent
1236	C	T	C	3rd	Silent
1284	A	A	G	3rd	Silent
1342	**A**	**G**	**G**	**1st**	**Domestic pig**
1427	T	T	C	2nd	Babirusa
1453	**T**	**C**	**C**	**1st**	**Domestic pig**
1591	**T**	**C**	**T**	**1st**	**Warthog**
1599	C	C	T	3rd	Silent

B

Amino Acid Position	Corresponding codon	Domestic pig	Warthog	Babirusa
131	391-393	Asparagine (N)	Asparagine (N)	Serine (S)
221	661-663	Lysine (K)	Lysine (K)	Glutamic acid (E)
258	772-774	Alanine (A)	Alanine (A)	Serine (S)
378	1132-1134	Serine (S)	Serine (S)	Alanine (A)
381	1141-1143	Alanine (A)	Alanine (A)	Valine (V)
448	1342-1344	**Threonine (T)**	**Alanine (A)**	Alanine (A)
476	1426-1428	Phenylalanine (F)	Phenylalanine (F)	Serine (S)
485	1453-1455	**Serine (S)**	**Proline (P)**	Proline (P)
531	1591-1593	**Serine (S)**	**Proline (P)**	Serine (S)

To confirm these findings, PCR primers were designed to amplify a 268bp region encompassing all three warthog/domestic pig differences directly from DNA: Fwd 5'-GGA AGG GAC ACT GAC AGA GG-3' and Rev 5'-TCA GAA GGG CTG AGA AGT CC-3'. This region was amplified and sequenced in a further fourteen warthogs, eight domestic pigs and six bushpigs (*Potamochoerus spp.*). These additional sequencing results confirmed the initial results, with domestic pig and warthog alleles being consistent in all samples examined. The inclusion of sequence from ASFV-resistant bushpigs indicates that residues 448 and 485 are conserved between warthogs and bushpigs. However, the presence of a proline residue at position 531 appears unique to warthogs. Bushpigs have a serine residue at this position 531 in common with domestic pig, babirusa and human (Figure 3.9).

3.7 TNFα Promoter

The most functionally important region of a eukaryotic polymerase II promoter can be considered to be the well-conserved ~500bp immediately preceding the transcription start point (Scherf *et al.* 2000). However, as the true extent of a promoter is difficult to define, the entire TNF 'intergenic region' was amplified. This region comprises approximately 1.6kbp of sequence which lies between the gene encoding TNFβ and the coding sequence of TNFα. Therefore, for the purposes of this study, the 'TNFα promoter' was taken from the first base after the TNFβ polyadenylation signal to the base preceding the start codon of TNFα. This sequence is 1600bp in length in domestic pig, 1592bp in warthog and 1599bp in babirusa. The equivalent human sequence is 1503bp in length (Figure 3.10). Suid and human TNFα promoter regions differ through numerous nucleotide differences and several short insertion/deletion events spread throughout the sequence. Between warthog and domestic pig TNFα promoter sequence there are 30 nucleotide differences. In addition, the warthog TNFα promoter sequence possesses two single base deletions and one six base pair deletion with respect to the domestic pig (Figure 3.11). In general, these differences appear to be dispersed fairly evenly across the promoter.

Figure 3.9 – *p65* Polymorphic Region Comparative Sequence Analysis

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
PL	*Potamochoerus larvatus*	African bushpig
PP	*Potamochoerus porcus*	Red river hog

A Partial p65 nucleotide sequence (268bp) encompassing the transactivation domains, within which the 3 polymorphisms were seen. This sequence is numbered as the full p65 sequence in Figure 3.8.

B Predicted primary protein structure (90 amino acids) of the polymorphic region of p65, as numbered in Figure 22b. Residues 448 and 485 are conserved between warthogs and bushpigs; however the presence of a proline residue at position 531 is only seen in warthogs. The bushpigs (*Potamochoerus spp.*) encode serine residue at position 531, in common with the domestic pig and human sequence.

A

```
SS1   cag ttt gat act gat gag gac ctg ggg gcc ctg ctc ggc aat aac act gac ccg acc gtg  [1392]
SS2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
SS3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
SS4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
SS5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
SS6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
SS7   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
SS8   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA1   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA2   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA3   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA4   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA5   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA6   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA7   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA8   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA9   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA10  ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA11  ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA12  ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA13  ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PA14  ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PL1   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PL2   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PL3   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PL4   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PL5   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]
PP1   ... ... ... g.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1392]

SS1   ttc acg gac ctg gca tcc gtc gac aac tct gag ttt cag cag ctg ctg aac cag ggt gta  [1452]
SS2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
SS3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
SS4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
SS5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
SS6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
SS7   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
SS8   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA1   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA7   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA8   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA9   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA10  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA11  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA12  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA13  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PA14  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PL1   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PL2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PL3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PL4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PL5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
PP1   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1452]
```

```
SS1   tcc atg ccc ccc cac aca gct gag ccc atg ctg atg gag tac cct gag gct ata act cgc  [1512]
SS2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
SS3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
SS4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
SS5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
SS6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
SS7   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
SS8   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA1   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA2   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA3   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA4   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA5   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA6   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA7   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA8   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA9   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA10  c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA11  c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA12  c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA13  c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PA14  c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PL1   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PL2   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PL3   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PL4   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PL5   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]
PP1   c.. ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1512]


SS1   ttg gtg aca ggg tcc cag aga ccc cct gac cca gct ccc act ccc ctg ggg gcc tct ggg  [1572]
SS2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
SS3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
SS4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
SS5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
SS6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
SS7   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
SS8   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA1   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA6   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA7   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA8   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA9   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA10  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA11  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA12  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA13  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PA14  ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PL1   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PL2   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PL3   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PL4   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PL5   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
PP1   ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ... ...  [1572]
```

```
SS1  ctc acc aac ggt ctc ctc tcg ggg gac g [1600]
SS2  ... ... ... ... ... ... ... ... ... . [1600]
SS3  ... ... ... ... ... ... ... ... ... . [1600]
SS4  ... ... ... ... ... ... ... ... ... . [1600]
SS5  ... ... ... ... ... ... ... ... ... . [1600]
SS6  ... ... ... ... ... ... ... ... ... . [1600]
SS7  ... ... ... ... ... ... ... ... ... . [1600]
SS8  ... ... ... ... ... ... ... ... ... . [1600]
PA1  ... ... ... ... ... ... c.. ... ... . [1600]
PA2  ... ... ... ... ... ... c.. ... ... . [1600]
PA3  ... ... ... ... ... ... c.. ... ... . [1600]
PA4  ... ... ... ... ... ... c.. ... ... . [1600]
PA5  ... ... ... ... ... ... c.. ... ... . [1600]
PA6  ... ... ... ... ... ... c.. ... ... . [1600]
PA7  ... ... ... ... ... ... c.. ... ... . [1600]
PA8  ... ... ... ... ... ... c.. ... ... . [1600]
PA9  ... ... ... ... ... ... c.. ... ... . [1600]
PA10 ... ... ... ... ... ... c.. ... ... . [1600]
PA11 ... ... ... ... ... ... c.. ... ... . [1600]
PA12 ... ... ... ... ... ... c.. ... ... . [1600]
PA13 ... ... ... ... ... ... c.. ... ... . [1600]
PA14 ... ... ... ... ... ... c.. ... ... . [1600]
PL1  ... ... ... ... ... ... ... ... ... . [1600]
PL2  ... ... ... ... ... ... ... ... ... . [1600]
PL3  ... ... ... ... ... ... ... ... ... . [1600]
PL4  ... ... ... ... ... ... ... ... ... . [1600]
PL5  ... ... ... ... ... ... ... ... ... . [1600]
PP1  ... ... ... ... ... ... ... ... ... . [1600]
```

B

```
SS1   QFDTDEDLGA LLGNNTDPTV FTDLASVDNS EFQQLLNQGV SMPPHTAEPM LMEYPEAITR LVTGSQRPPD  [514]
SS2   .......... .......... .......... .......... .......... .......... ..........  [514]
SS3   .......... .......... .......... .......... .......... .......... ..........  [514]
SS4   .......... .......... .......... .......... .......... .......... ..........  [514]
SS5   .......... .......... .......... .......... .......... .......... ..........  [514]
SS6   .......... .......... .......... .......... .......... .......... ..........  [514]
SS7   .......... .......... .......... .......... .......... .......... ..........  [514]
SS8   .......... .......... .......... .......... .......... .......... ..........  [514]
PA1   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA2   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA3   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA4   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA5   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA6   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA7   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA8   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA9   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA10  ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA11  ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA12  ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA13  ...A...... .......... .......... .......... P......... .......... ..........  [514]
PA14  ...A...... .......... .......... .......... P......... .......... ..........  [514]
PL1   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PL2   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PL3   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PL4   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PL5   ...A...... .......... .......... .......... P......... .......... ..........  [514]
PP1   ...A...... .......... .......... .......... P......... .......... ..........  [514]
```

```
SS1   PAPTPLGASG LTNGLLSGD [533]
SS2   .......... ......... [533]
SS3   .......... ......... [533]
SS4   .......... ......... [533]
SS5   .......... ......... [533]
SS6   .......... ......... [533]
SS7   .......... ......... [533]
SS8   .......... ......... [533]
PA1   .......... ......P.. [533]
PA2   .......... ......P.. [533]
PA3   .......... ......P.. [533]
PA4   .......... ......P.. [533]
PA5   .......... ......P.. [533]
PA6   .......... ......P.. [533]
PA7   .......... ......P.. [533]
PA8   .......... ......P.. [533]
PA9   .......... ......P.. [533]
PA10  .......... ......P.. [533]
PA11  .......... ......P.. [533]
PA12  .......... ......P.. [533]
PA13  .......... ......P.. [533]
PA14  .......... ......P.. [533]
PL1   .......... ......... [533]
PL2   .......... ......... [533]
PL3   .......... ......... [533]
PL4   .......... ......... [533]
PL5   .......... ......... [533]
PP1   .......... ......... [533]
```

Figure 3.10 – TNFα Promoter Comparative Sequence Analysis

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
BB	*Babyrousa babyrussa*	Babirusa
HS	*Homo sapiens*	Human (M16441)

TNFα promoter sequence, which includes the 5' untranslated region, was taken as the first base after the TNFβ polyadenylation signal to the base preceding the start codon of TNFα. The TNFα promoter is 1600bp in length in domestic pig, 1592bp in the warthog and 1599 bp in the babirusa. The human sequence is 1503bp in length. All pig sequences contain three copies of a conserved 13bp tandem repeat (dotted underlined), of which the human sequence contains only one. The TATA box has been underlined and the transcription start point highlighted (C); both are conserved between all four species.

```
SS GGCTCAGGAA GG-GGCTGCT TGACTGGAGG CTCATGAGGA GACGGCTG-A CCCTCGATGA AACCCAATAA  [  70]
PA .......... ..-....... .......... .......... ........-. .......... ..........  [  70]
BB .......... ..-....... .......... .......... ...A....-. .......... ..........  [  70]
HS .....T.A.. .CCA.....C -...CA...C .C..CAC... .G.AT...C. .......... .G........  [  70]

SS AGCTCTTTTC TCTGAAATGC TGTCTGCTCG TATCTGTCAC TCGGGAGGGG AGAATTCTCC AGATGTCTCT  [ 140]
PA .......... .......... .......C.. .......... .......... .......... ..........  [ 140]
BB .......... .......... .......C.. .......... .......... .......... ..........  [ 140]
HS .C........ .......... ........T. .G.G...GTG ..T..GA.T. ....C.TC.. ..-.C.A...  [ 140]

SS AAGGAGTGGA GGGAGG---A CAGGAATCAG AGGGGACGGG AGCTGTGGGT G---TGTGAT GA----GGCC  [ 210]
PA .......... ......---. .......... .......A.. .........C A---.....A ..------....  [ 210]
BB .......... ......---. .......... .......... .........C .---..CA.A ..------....  [ 210]
HS .....A.... ......GAC. G...GC...A ....AG.AA. .........G AGAACAAA.G ..TAAG...T  [ 210]

SS TAAGGGGCTC AGGTGAGAGA TGGCGGCCTC A--GGGTGAG GGCAGCCAGA CCCCTGCAGG AGAAGCAGAT  [ 280]
PA .......... .......... .....A.... G--........ .A........ ....-..... ..........  [ 280]
BB .......... .......... .......... G--........ .......... .......... ..........  [ 280]
HS C.GA.A...T CA.G..T.TG ..AT..A... .CCA...... .-.C...... .TG....... G......A.G  [ 280]

SS GGTCCTCTGA GAAGACAAAG GAAGAGATGC AGGGCCAAGG TCTTGAGAAC CGAGG-TCGG GGGTCGCCTG  [ 350]
PA .......... .......... .......... .......... ...G...... .....-..A. ..........  [ 350]
BB ........C. .......... ..-....... .......... ...G...... .....-..A. ..........  [ 350]
HS .AGAAG.... .....TG... ..-------- .AA.T..G.. ...G...GGG ..G..G..A. ..AG.T....  [ 350]

SS GCAGATATGG CCACAGGTAG AGGGACAGAG GAATAGGGGT GACAGGAGGC TTCCCGGGAG AAGGGAACAC  [ 420]
PA .......... .......... ......G... .......... .......... .......A.. C.........  [ 420]
BB .......C.. .......... ......G... .......... .......... .......... C.........  [ 420]
HS .G........ .....T.... C..CT.T... ....--.... T.......A. C..TG..... .T.T..C...  [ 420]

SS ACTGAGGGGT GTTCGGGATT CTGAGGGAGG AGCACGGGGA CGCCCTGGGA GACATGGCCG TGCCAGGGCC  [ 490]
PA .......... ...T...... .......... .......... .......... .........A ..........  [ 490]
BB .......... .......... .......... .......... T......... T.T....... ..........  [ 490]
HS .GCA.T.... ---------- -------... ..------A. T.T..A...- --------.TA ..--.AA.T.  [ 490]

SS ATGAGGAGTG GGAGAGCCTC TGAGGCTAGC GGCTGGAGAT ACAGGGACAT TTGAGGAGAC ACGGTCATGG  [ 560]
PA C......... .......... .......... T......... .......... .......... .......---  [ 560]
BB C......... .......... .......... .......... .......... .......... .......C.C  [ 560]
HS -----....A T.G.GA..C. CCC--T..A. ----.A..-- ......C... G.AGA.G.C. C.A.------  [ 560]

SS CCAGGAGCGC CGAGGGCCTG GACAGTCTCT AGGAATCTCG AAGAAGCAGG AATTCTTTGA GGATACGTGG  [ 630]
PA ---....... .......... .G........ ........G. .......... .......... ..........  [ 630]
BB ........A. T......... .........C ........G. .......... .......... .....T....  [ 630]
HS ---....T.A -A..A....C C.GGAC...C ...T..--G. ..T.--.... GGACG...A. .A.G.TA...  [ 630]

SS CCACACAAAG GGAGGCTGAG GTGTGGGGAC TTCATGCAGA AG-TCAGGG- CCTCACATTC CCTTGGAAGC  [ 700]
PA .......... .......... .......... .......... ..-.......- ..C....... ..........  [ 700]
BB .......... ...A...... T......... .......... ..-.......- .......... ..........  [ 700]
HS .......CT. ..GCC..... AA...A.AG. ......A.A. .AA.......A ..C..G.G.T ..........  [ 700]

SS CGAGACTGAA ACCAGCAGCA GAGTTTTGGT GAGTTCCTGT CAGAGTGAAA GGAGAAGGCC CGCCATGGTG  [ 770]
PA .......... .......... .......... .......... .......... .......... ..........  [ 770]
BB A......... .......... .......... .......... .......... .......... ..........  [ 770]
HS .A........ .........--- ---..A..A- .TC.C.GG.. ....A..... .A...G.... T...CCA...  [ 770]

SS GG-TTTGTGA ATTCCCAGCC TGGCTTCCTC TCCCTCTGGG GCTGTCCCAG GCCTGTTCCT GCCGTCCTCC  [ 840]
PA ..-....... .......... .......... .......... .......... .......... ..........  [ 840]
BB ..-....... .......... .......... .......... .......... .......... ..........  [ 840]
HS ..G.C..... ......G.G- -..TGATT.. A.T.C.C... .......... ..T...C... ..TAC..G.A  [ 840]

SS CCCAGCCCGT G-TAGGGCCT CCAGC-TGCC CTT---CTCC CAGCTCCTCT TCCCTCCAGG AGACGAAACA  [ 910]
PA .A........ .-........ .....-.... ...---.... .......... .......... ..........  [ 910]
BB .......... .-........ .....-.... ...---.... .......... .......... ..........  [ 910]
HS .......TT. CC.GA..... .A...C.... ACCAAG.C.. ........TC ....CG.... -.C.C.....  [ 910]

SS TGGGTCTCAG CACCCAGCGC GGTGT---CG TCTAA----G TTTTCTCTCC ATTAA-GAAC TCAGCTTTCT  [ 980]
PA .......... .......... .....---.. .....----. .......... .....-.... ..........  [ 980]
BB .......... .......... .....---.. .....----. .......... .....-.G.. ..........  [ 980]
HS CA..C..... G..T..A.A. A.CT.TTC.C ..C..CCCC. .......... C.C..C.G.. ..........  [ 980]
```

```
SS GAAGCTCCTC CCATTCCTAG TTCTACCCCT A-CCTGAGCC CTGTTCGGAA ATCAGAGAGA AATAGAAGTC  [1050]
PA .......... .......... .......... .-........ ......A... .........T ..........  [1050]
BB .......... .......... .......... .-........ .......... .......... ..........  [1050]
HS .....C.... ...G.T.... .....T.TT. TT....CAT. ....CT.... G.T...AG.. ..C...----  [1050]

SS ATCCCCCAAA GAAAAGGAAT TTGTCCCCCA AAGAAACAGA ACTTGTCCCC CAAAGAAATG GAAACAATGG  [1120]
PA .......... .......... .......... .......... .......... .......... ..........  [1120]
BB .......... .......... .......... .......... .......... .......... ..........  [1120]
HS -----...-- ---------- ---------- ------.... C..G...... A......... ..GG....A.  [1120]

SS GAAATGGGAG GCAGGGGGGA CCTGGGGTCC AGCCTCCAGG GTCCTACACA CAGAGCAGTA ACTGGCCCAG  [1190]
PA .......... ..-....... .......... .......... .........G ......G... ..........  [1190]
BB .......... ........C. A......... .......... .......... .......... ..........  [1190]
HS .TTT..A.G. ...T....A- --C.....T. .......... .......... ..A.T....C .G........  [1190]

SS CAAGCCCACC TCAGGATCCG GGCAGGGAGG GTAGGAAGT- --------AT CCCTGATGCC TGGGTGTCCC  [1260]
PA .......... .......... ...G...... .........- --------.. .......... ..........  [1260]
BB .......... .......... ...G...... .........- --------.. .......... ..........  [1260]
HS A.GA...C.. ..G.A...G. A......... A.G..G...G TGAGGGGT.. ..T......T ..T.......  [1260]

SS CAACTTTCCA AACCGCCGCC CCCGCTATGG AGATGAAACT AAGACAGAAG GTGCAGGGCC CGCTACCGCT  [1330]
PA .......... .......A.. .......... .......... G......... .......... ....G.....  [1330]
BB .......... .......... .......... .......... G......... .......... ..........  [1330]
HS .......... ..T.C..... .....G.... ...A.....C G......... .......... .A........  [1330]

SS TCCTCCAGAT GAGCTCATGG GTTTCTCCAC CAAGGAAGTT TTCCGCTGGT TGAAAGAGAG CCTCTCCCCG  [1400]
PA .......... .......... .......... .......... .......... .......... ..........  [1400]
BB .......... .......... .......... .......... .......... .......... ..........  [1400]
HS .......... .......... .......... .......... .......... ....T..--T T..T......  [1400]

SS CCCTCTTCTC ACCC--AGAG CGTATAAATG CAGCTGTTTG CACACCCAGC CAGCAGAAGC TCCCAGAGTG  [1470]
PA .......... ....--.... .......... .......... .......... .......... ..........  [1470]
BB .......... ....--.... .......... .......... .......... .......... ..........  [1470]
HS .....C.... G...CAG.GA .A......G. ...T....G. .......... .......C.. ....TC..CA  [1470]

SS AGGACACCAG GGGACCAGCC AGGAGAGAGA CAAGCCACTC CAGGACCCCC TAGAAATAAC CTCTCAGAAG  [1540]
PA .......... .......... .......... .......... .......... .......... ..........  [1540]
BB .......... .......... .......... .......... ...A...... .......... ..........  [1540]
HS ......G... A........T .A...G.... G....A...A ...AC..... CT....AC.A .C......C.  [1540]

SS ACACACCCCC GAACAGGCAG CCGGAC-GAC TCTCTCCCTC TCACACGCTG CCCCGGGGCG CCACCATCTC  [1610]
PA .......... .......... ......-... .......... .......... ....A....A ..........  [1610]
BB .......... .......... ......-... ...T.A.... .......... .......... ..........  [1610]
HS C....T.... TG...A..T. ..A.G.A.GT .....T.... .....TA... A...AC...T T....C....  [1610]

SS CCAGCTGGAC CTGAGCCCCT CTGAAAAAGA CACC [1644]
PA .......... .......... .......... .... [1644]
BB .......... .......... .......... .... [1644]
HS T.-------- --------.C ...G...G.. .... [1644]
```

Figure 3.11 – Domestic Pig versus Warthog TNFα Promoter Sequence

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog

The TNFα promoter sequences of the domestic pig (1600bp) and warthog (1592bp) differ at 30 positions. In addition, the warthog sequence displays two single base pair deletions and one six base pair deletion, with respect to the domestic pig. The TATA box has been underlined and the transcription start point highlighted (C). All sequence downstream from this point (1419bp) is transcribed to become the 5' untranslated region of the TNFα mRNA transcript. Both sequences contain three copies of a conserved 13bp tandem repeat (dotted underlined).

```
SS GGCTCAGGAA GGGGCTGCTT GACTGGAGGC TCATGAGGAG ACGGCTGACC CTCGATGAAA CCCAATAAAG  [  70]
PA .......... .......... .......... .......... .......... .......... ..........  [  70]

SS CTCTTTTCTC TGAAATGCTG TCTGCTCGTA TCTGTCACTC GGGAGGGGAG AATTCTCCAG ATGTCTCTAA  [ 140]
PA .......... .......... .....C.... .......... .......... .......... ..........  [ 140]

SS GGAGTGGAGG GAGGACAGGA ATCAGAGGGG ACGGGAGCTG TGGGTGTGTG ATGAGGCCTA AGGGGCTCAG  [ 210]
PA .......... .......... .......... ..A....... ....CA.... .A........ ..........  [ 210]

SS GTGAGAGATG GCGGCCTCAG GGTGAGGGCA GCCAGACCCC TGCAGGAGAA GCAGATGGTC CTCTGAGAAG  [ 280]
PA .......... ...A....G. .......A.. .......... -......... .......... ..........  [ 280]

SS ACAAAGGAAG AGATGCAGGG CCAAGGTCTT GAGAACCGAG GTCGGGGGTC GCCTGGCAGA TATGGCCACA  [ 350]
PA .......... .......... .........G .......... ...A...... .......... ..........  [ 350]

SS GGTAGAGGGA CAGAGGAATA GGGGTGACAG GAGGCTTCCC GGGAGAAGGG AACACACTGA GGGGTGTTCG  [ 420]
PA .......... .G........ .......... .......... ..A..C.... .......... ........T.  [ 420]

SS GGATTCTGAG GGAGGAGCAC GGGGACGCCC TGGGAGACAT GGCCGTGCCA GGGCCATGAG GAGTGGGAGA  [ 490]
PA .......... .......... .......... .......... ....A..... .....C.... ..........  [ 490]

SS GCCTCTGAGG CTAGCGGCTG GAGATACAGG GACATTTGAG GAGACACGGT CATGGCCAGG AGCGCCGAGG  [ 560]
PA .......... .....T.... .......... .......... .......... ..------.. ..........  [ 560]

SS GCCTGGACAG TCTCTAGGAA TCTCGAAGAA GCAGGAATTC TTTGAGGATA CGTGGCCACA CAAAGGGAGG  [ 630]
PA ......G... .......... ...G...... .......... .......... .......... ..........  [ 630]

SS CTGAGGTGTG GGGACTTCAT GCAGAAGTCA GGGCCTCACA TTCCCTTGGA AGCCGAGACT GAAACCAGCA  [ 700]
PA .......... .......... .......... .....C.... .......... .......... ..........  [ 700]

SS GCAGAGTTTT GGTGAGTTCC TGTCAGAGTG AAAGGAGAAG GCCCGCCATG GTGGGTTTGT GAATTCCCAG  [ 770]
PA .......... .......... .......... .......... .......... .......... ..........  [ 770]

SS CCTGGCTTCC TCTCCCTCTG GGGCTGTCCC AGGCCTGTTC CTGCCGTCCT CCCCCAGCCC GTGTAGGGCC  [ 840]
PA .......... .......... .......... .......... .......... ...A...... ..........  [ 840]

SS TCCAGCTGCC CTTCTCCCAG CTCCTCTTCC CTCCAGGAGA CGAAACATGG GTCTCAGCAC CCAGCGCGGT  [ 910]
PA .......... .......... .......... .......... .......... .......... ..........  [ 910]

SS GTCGTCTAAG TTTTCTCTCC ATTAAGAACT CAGCTTTCTG AAGCTCCTCC CATTCCTAGT TCTACCCCTA  [ 980]
PA .......... .......... .......... .......... .......... .......... ..........  [ 980]

SS CCTGAGCCCT GTTCGGAAAT CAGAGAGAAA TAGAAGTCAT CCCCCAAAGA AAAGGAATTT GTCCCCCAAA  [1050]
PA .......... ....A..... .......T.. .......... .......... .......... ..........  [1050]

SS GAAACAGAAC TTGTCCCCCA AAGAAATGGA AACAATGGGA AATGGGAGGC AGGGGGGACC TGGGGTCCAG  [1120]
PA .......... .......... .......... .......... .......... -......... ..........  [1120]

SS CCTCCAGGGT CCTACACACA GAGCAGTAAC TGGCCCAGCA AGCCCACCTC AGGATCCGGG CAGGGAGGGT  [1190]
PA .......... .......G.. ....G..... .......... .......... .......... .G........  [1190]

SS AGGAAGTATC CCTGATGCCT GGGTGTCCCC AACTTTCCAA ACCGCCGCCC CCGCTATGGA GATGAAACTA  [1260]
PA .......... .......... .......... .......... ......A... .......... .........G  [1260]

SS AGACAGAAGG TGCAGGGCCC GCTACCGCTT CCTCCAGATG AGCTCATGGG TTTCTCCACC AAGGAAGTTT  [1330]
PA .......... .......... ...G...... .......... .......... .......... ..........  [1330]

SS TCCGCTGGTT GAAAGAGAGC CTCTCCCCGC CCTCTTCTCA CCCAGAGCGT ATAAATGCAG CTGTTTGCAC  [1400]
PA .......... .......... .......... .......... .......... .......... ..........  [1400]

SS ACCCAGCCAG CAGAAGCTCC CAGAGTGAGG ACACCAGGGG ACCAGCCAGG AGAGAGACAA GCCACTCCAG  [1470]
PA .......... .......... .......... .......... .......... .......... ..........  [1470]

SS GACCCCCTAG AAATAACCTC TCAGAAGACA CACCCCCGAA CAGGCAGCCG GACGACTCTC TCCCTCTCAC  [1540]
PA .......... .......... .......... .......... .......... .......... ..........  [1540]

SS ACGCTGCCCC GGGGCGCCAC CATCTCCCAG CTGGACCTGA GCCCCTCTGA AAAAGACACC  [1600]
PA .......... A....A.... .......... .......... .......... ..........  [1600]
```

3.7.1 Predicted Transcription Factor Binding Domains.

Possible differences in potential transcription factor binding domains were analysed in the TNFα promoter sequences of all species studied here (Genomatix software package; http://www.genomatix.de). MatInspector™ (software release 7.2.2 April 2004) was initially utilised to predict individual transcription factor binding domains (from a vertebrate library of 1188, divided into 262 families) (Quandt *et al.* 1995). The quality of these predictions is assessed by a number of calculations which result in a value between 0 and 1. The most useful guide is 'Matrix Similarity', which takes into account how highly conserved each position within a characterised binding domain is, and then allocates a score to the query sequence depending on this. For example, it strongly rewards a match at a highly conserved position, but does not heavily penalise a mismatch at a position which is poorly conserved (and vice versa). The results were restricted to those with a matrix similarity greater than 0.75 as advised by Klaus May (Genomatix, personal communication).

Using this software, it was immediately apparent that TNFα promoter is rich in putative transcription factor binding domains, including those for NFκB, NFAT, activator protein (AP)-1, AP-2, AP-4, stimulating protein (SP)-1, serum response factor (SRF), interferon regulatory factor (IRF)-3, IRF-7, CAAT enhancer-binding protein (C/EBP), tumour suppressor p53, signal transducer and activator of transcription (STAT)-5, STAT-6, E2F, Smad 3, Smad 4 and octamer-binding factor (OBF)-1 sites. MatInspector predicted 170 potential binding sites in the domestic pig (Appendix 1), 173 in the warthog (Appendix 2) and 178 in the babirusa. The human sequence contained 156 matches. A vast majority of these predicted transcription factor binding sites (149) are conserved between warthogs and domestic pigs. However, the MatInspector software highlighted a number of binding sites as being present in only one or other of the species; there are 21 sites present in only the domestic pig sequence (Table 3.8) and 24 sites unique to warthog (Table 3.9).

Of the unique sites present, the six base pair deletion event in the warthog removes a putative p53 half-site, which is present on both the positive and negative strands in the domestic pig. However, the first single point deletion in the warthog sequence

Table 3.8

The 21 transcription factor binding sites, predicted by MatInspector, which are unique to the domestic pig sequence. MatInspector provides a code to describe the family of each transcription factor, a brief description of the individual transcription factor, the location of the binding domain, which DNA strand (+ or -) the domain is encoded by, the matrix similarity and the nucleotide sequence of the domain. Uppercase letters denote the four most highly conserved bases (core) within the domain.

Family/matrix	Description	Position	Str.	Matrix sim.	Sequence
V$EVI1/EVI1.04	Ecotropic viral integration site 1 encoded factor	87 - 103	(-)	0.803	aGATAcgagcagacagc
V$AP4R/AP4.03	Activator protein 4	169 - 185	(-)	0.863	acccaCAGCtcccgtcc
V$RREB/RREB1.01	Ras-responsive element binding protein 1	171 - 185	(-)	0.791	aCCCAcagctcccgt
V$MINI/MUSCLE_INI.02	Muscle Initiator Sequence	223 - 241	(-)	0.867	ctgcccTCACcctgaggcc
V$DEAF/NUDR.01	NUDR (nuclear DEAF-1 related transcriptional regulator protein)	319 - 337	(+)	0.732	aggTCGGgggtcgcctggc
V$NOLF/OLF1.01	Olfactory neuron-specific factor	381 - 403	(+)	0.823	gaggctTCCCgggagaagggaac
V$E2FF/E2F.01	E2F, involved in cell cycle regulation, interacts with Rb p107 protein	386 - 400	(+)	0.786	ttcccggGAGAaggg
V$DEAF/NUDR.01	NUDR (nuclear DEAF-1 related transcriptional regulator protein)	415 - 433	(+)	0.757	tgtTCGGgattctgaggga
V$P53F/P53.01	Tumor suppressor p53	455 - 475	(+)	0.687	agaCATGgccgtgccagggcc
V$HAND/HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12	467 - 481	(+)	0.775	gccaggGCCAtgagg
V$P53F/P53.02	Tumor suppressor p53 (5' half site)	528 - 548	(+)	0.910	gaggagacacggtCATGgcca
		537 - 557	(-)	0.912	cggcgctcctggcCATGaccg
V$ARP1/ARP1.01	apolipoprotein AI regulatory protein 1	652 - 668	(-)	0.826	tgaggcCCTGacttctg
V$ZBPF/ZBP89.01	Zinc finger transcription factor ZBP-89	811 - 833	(+)	0.961	ctgccgtcctCCCCcagcccgtg
V$AP2F/AP2.01	Activator protein 2	819 - 831	(+)	0.890	ctCCCCcagcccg
V$IRFF/IRF7.01	Interferon regulatory factor 7 (IRF-7)	1005 - 1023	(+)	0.871	gaGAAAtagaagtcatccc
V$VMYB/VMYB.04	v-Myb, AMV v-myb	1145 - 1155	(+)	0.852	agtAACTggcc
V$CMYB/CMYB.02	c-Myb, important in hematopoesis, cellular equivalent to avian myoblastosis virus oncogene v-myb	1145 - 1157	(+)	0.966	agTAACtggccca
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	1230 - 1244	(-)	0.919	gcgggGGCGgcggtt
V$AP2F/AP2.01	Activator protein 2	1544 - 1556	(-)	0.907	cgCCCCggggcag
		1545 - 1557	(+)	0.915	tgCCCCggggcgc

Table 3.9

The 24 transcription factor binding sites, predicted by MatInspector, which are unique to the warthog sequence. MatInspector provides a code to describe the family of each transcription factor, a brief description of the individual transcription factor, the location of the binding domain, which DNA strand (+ or -) the domain is encoded by, the matrix similarity and the nucleotide sequence of the domain. Uppercase letters denote the four most highly conserved bases (core) within the domain.

Family/matrix	Description	Position	Str.	Matrix sim.	Sequence
V$P53F/P53.02	Tumor suppressor p53 (5' half site)	181 - 201	(-)	0.924	ttaggcctcttcaCATGccca
V$BNCF/BNC.01	Basonuclin, cooperates with USF1 in rDNA PolI transcription)	232 - 250	(-)	0.880	ggggtctggcTGTCctcac
V$GREF/GRE.01	Glucocorticoid receptor, C2C2 zinc finger protein binds glucocorticoid dependent to GREs	233 - 251	(-)	0.890	cggggtctggctGTCCtca
V$AP2F/AP2.01	Activator protein 2	246 - 258	(+)	0.897	acCCCGcaggaga
V$RORA/RORA1.01	RAR-related orphan receptor alpha1	312 - 328	(+)	0.936	gaaccgaGGTCaggggt
V$ETSF/CETS1P54.01	c-Ets-1(p54)	384 - 400	(+)	0.973	cttccCGGAagcaggga
V$IKRS/IK3.01	Ikaros 3, potential regulator of lymphocyte differentiation	393 - 405	(+)	0.857	agcagGGAAcaca
V$RREB/RREB1.01	Ras-responsive element binding protein 1	408 - 422	(-)	0.879	tCCCAaacacccctc
V$IKRS/IK2.01	Ikaros 2, potential regulator of lymphocyte differentiation	415 - 427	(+)	0.983	gtttGGGAttctg
V$MOKF/MOK2.01	Ribonucleoprotein associated zinc finger protein MOK-2 (mouse)	465 - 485	(-)	0.761	ccactcctcagggCCCTggca
V$NOLF/OLF1.01	Olfactory neuron-specific factor	465 - 487	(-)	0.836	tcccacTCCTcagggccctggca
V$AP4R/TH1E47.01	Thing1/E47 heterodimer, TH1 bHLH member specific expression in a variety of embryonic tissues	569 - 585	(-)	0.932	gcttcttCCAGattcct
V$HOXT/MEIS1_HOXA9.01	Homeobox protein MEIS1 binding site	984 - 996	(-)	0.824	cTGATttctgaac
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	1085 - 1097	(+)	0.910	atggGAGGcgggg
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	1086 - 1100	(+)	0.899	tgggaGGCGgggga
V$MAZF/MAZR.01	MYC-associated zinc finger protein related transcription factor	1090 - 1102	(+)	0.883	aggcggGGGGacc
V$RXRF/FXRE.01	Farnesoid X - activated receptor (RXR/FXR dimer)	1136 - 1152	(+)	0.808	cggtaacTGGCccagca
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	1166 - 1180	(+)	0.920	atccgGGCGgggagg
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	1168 - 1180	(+)	0.904	ccggGCGGggagg
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	1172 - 1178	(+)	0.985	gcGGGGa
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	1173 - 1185	(+)	0.911	cgggGAGGgtagg
V$EBOX/ATF6.01	Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y	1224 - 1238	(+)	0.937	ccgCCACccccgcta
V$IRFF/IRF3.01	Interferon regulatory factor 3 (IRF-3)	1244 - 1262	(+)	0.860	atgaaactGAGAcagaagg
V$AP2F/AP2.01	Activator protein 2	1268 - 1280	(+)	0.910	ggCCCGctgccgc

introduces a putative activator protein 2 site and the second deletion introduces a zinc finger transcription factor binding site (a myc-associated zinc finger protein transcription factor and/or stimulating protein 1). None of the candidate binding sites highlighted by MatInspector as being unique were solely represented in the sequence. Indeed, the software was actually detecting extra copies of binding sites already present in multiple locations within the promoter.

Although it is tempting to speculate on the possible significance of differences in individual predicted transcription factor binding sites, it is impossible to prove functionality without experimental data. Moreover, the importance of the precise location and copy number of individual transcription factor binding domains is difficult to predict. These factors, coupled with the appearance of false positive results (especially those with a short and/or repetitive core recognition sequence, for example GGGG), mean that such computer-generated data must be interpreted with caution. However, used as an initial analysis, these data suggest that differences in transcriptional activity of TNFα may exist between these suid species.

3.7.2 Predicted Transcription Factor Modules

A more stringent and robust method of predicting functional transcription factor binding domains is to look for transcription factor *modules*. Transcription factor modules comprise well-characterised, functional clusters of two or more transcription factor binding domains (Frech *et al.* 1997; Klingenhoff *et al.* 1999). Within a module, transcription factor binding domains occur in a particular order and within a certain distance from each other. As modules are larger, better characterised and more complex than single transcription factor binding domains, the risk of false positives is greatly reduced. As such, identification of species differences in transcription factor modules is more likely to be indicative of functional significance.

The domestic pig and warthog TNFα promoter sequences were entered into the Genomatix ModelInspector™ program (professional release 5.0.1, April 2004). This searched the sequences against the entire vertebrate module library (305 modules).

Results were restricted to those with a model score in excess of 80%. Model score is calculated by comparison with standard results achieved from a panel of characterised training sequences (Klaus May, Genomatix, personal communication).

Ten putative modules were identified in the domestic pig and seventeen in the warthog TNFα promoter sequence (Table 3.10). Only one module was unique to the domestic pig, compared with the warthog sequence. In comparison, the warthog contained eight modules which were not present in the domestic pig sequence. The location of most modules can be divided into two clear regions, proximal and distal (Figure 3.12). The proximal kilobase of sequence (immediately upstream of the transcription start point) is highly conserved between both species, encoding seven identical predicted modules. Within this region, domestic pig and warthog sequences also contain a similar eighth module, comprising an ETSF and SP1F motif, which may be equivalent. Within this proximal region, the warthog sequence encodes one unique module, SP1F_CEPB. In the distal promoter, a 63bp region of sequence between bases 358 and 421 contains two conserved overlapping modules in both species, NFKB_AP1F and ETSF_AP1F. However, within this relatively short region, the warthog sequence encodes a further five potential modules. Finally, in the region between the proximal and distal groups of modules, the warthog contains one extra module, BRNF_P53F, between bases 450 and 534.

The 63bp of distal sequence described above contains four base differences between the domestic pig and warthog. The central sequence containing the BRNF_P53F module in the warthog contains three base differences. The one extra module in the proximal region of the warthog sequence, SP1F_CEPB, is formed by a single base deletion which permits an SP1 site. Based on the high strength of the model scores, these seven base differences and single deletion event in warthog may represent key targets for future analysis.

Figure 3.12 – Transcription factor binding modules in the TNFα Promoter

Diagrammatic representation (not to scale) illustrating the location of putative transcription factor binding modules in the domestic pig and warthog TNFα promoter. The promoter is numbered from 0 (the base after the TNFβ polyadenylation signal) to 1600 (the base preceding the TNFα start codon). The promoter has been divided into proximal and distal regions as discussed in the main text.

Arrows indicate the DNA strand on which the module is encoded (→ positive strand; ← negative strand). Black arrows represent modules conserved between the warthog and domestic pig. Green arrows are modules present in only the warthog sequence. The single red arrow is the only module unique to the domestic pig. The two ETSF_SP1F modules marked with an asterisk (*) differ slightly in their conformation and precise location, but may be equivalent.

Putative Transcription Factor Modules on the TNFα Promoter

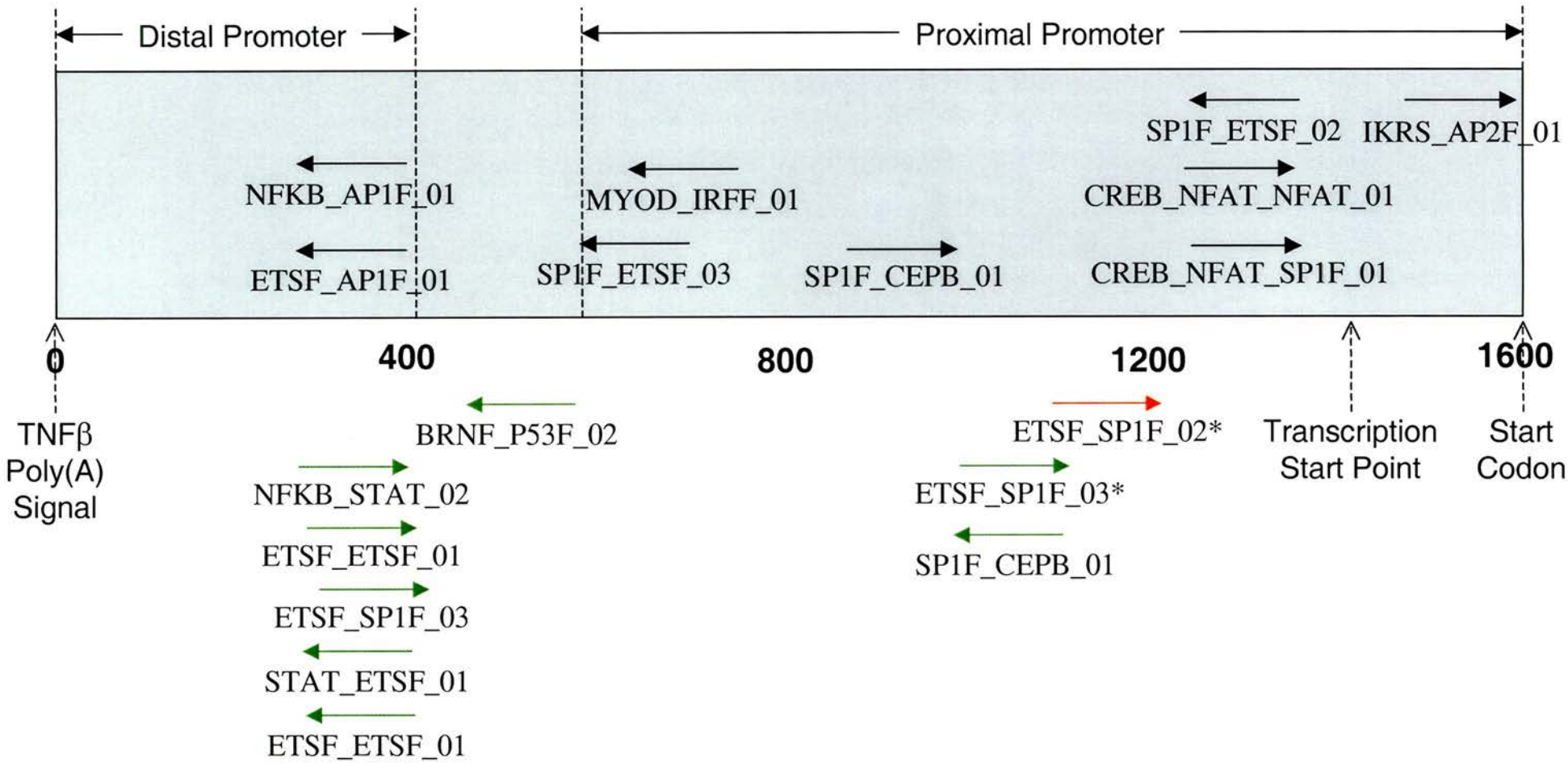


Table 3.10

The transcription factor binding domain modules identified by ModelInspector in the domestic pig and warthog TNFα promoter sequence. These sequences were searched against the entire vertebrate module library (305 modules). Ten modules were identified in the domestic pig (top) and seventeen in the warthog sequence (below). Only one module is unique to the domestic pig, compared to the warthog which contains an additional eight modules. Unique modules are highlighted in **bold**. Model score is calculated by comparison with standard results achieved from a panel of characterised training sequences; results were restricted to those with a model score in excess of 80%.

The location of most modules can be divided into two groups. Those within the highly conserved proximal kilobase of sequence (dark shading) and those within a 63bp region of distal sequence (light shading). Only one module, BRNF_P53F, lies between these regions and it is only present in the warthog sequence (intermediate shading).

Sequence	Module Name	Position	Strand	Model Score
Domestic Pig (1 - 1600)	NFKB_AP1F_01	392 - 373	(-)	88.3 %
	ETSF_AP1F_01	395 - 373	(-)	91.2 %
	SP1F_ETSF_03	635 - 588	(-)	88.0 %
	MYOD_IRFF_01	706 - 683	(-)	94.0 %
	SP1F_CEBP_01	951 - 974	(+)	81.8 %
	ETSF_SP1F_02	1167 - 1244	(+)	90.5 %
	CREB_NFAT_NFAT_01	1293 - 1336	(+)	84.7 %
	CREB_NFAT_SP1F_01	1293 - 1366	(+)	83.9 %
	SP1F_ETSF_02	1473 - 1282	(-)	89.4 %
	IKRS_AP2F_01	1526 - 1557	(+)	91.0 %

Sequence	Module Name	Position	Strand	Model Score
Warthog (1 - 1592)	**NFKB_STAT_02**	377 - 398	(+)	91.8 %
	ETSF_ETSF_01	378 - 400	(+)	93.1 %
	ETSF_SP1F_03	384 - 421	(+)	93.6 %
	NFKB_AP1F_01	391 - 372	(-)	88.3 %
	ETSF_AP1F_01	394 - 372	(-)	91.0 %
	STAT_ETSF_01	398 - 358	(-)	92.3 %
	ETSF_ETSF_01	400 - 378	(-)	94.4 %
	BRNF_P53F_02	534 - 450	(-)	82.4 %
	SP1F_ETSF_03	628 - 581	(-)	88.0 %
	MYOD_IRFF_01	699 - 676	(-)	94.0 %
	SP1F_CEBP_01	944 - 967	(+)	81.8 %
	ETSF_SP1F_03	1065 - 1100	(+)	87.6 %
	SP1F_CEBP_01	1100 - 1074	(-)	85.7 %
	CREB_NFAT_NFAT_01	1285 - 1328	(+)	84.7 %
	CREB_NFAT_SP1F_01	1285 - 1358	(+)	83.9 %
	SP1F_ETSF_02	1465 - 1274	(-)	89.2 %
	IKRS_AP2F_01	1518 - 1549	(+)	88.5 %

3.7.3 Phylogenetic Analysis

In addition to a domestic pig, warthog and babirusa, TNFα promoter sequence was also determined for two bearded pigs (*Sus barbatus subspp.*), an additional babirusa, two further warthogs, two African bushpigs (*Potamochoerus africanus*), two red river hogs (*Potamochoerus porcus*), a Javan warty pig (*Sus verrrucosus*), three Sulawesi warty pigs (*Sus celebensis*) and a white-lipped peccary (*Tayassu peccari*). One human and two extra domestic pig sequences were also retrieved from the public sequence database (Appendix 3). Although it does not form part of the remit of this thesis, these sequences have been combined with sequencing data from three other studies (microsatellite, *c-Kit* intron seven and *cytochrome b*) in order to produce a comprehensive molecular phylogeny of the suids. Together, these studies represent a total of ~4kbp of sequence from each animal, compiled from four distinct sequence types: promoter, microsatellite, intronic and mitochondrial. The sequencing of a panel of 26 microsatellites was carried out by Lowden *et al.* (2002), *c-Kit* intron seven sequencing was completed by Dr. Wee Tek Tay (Roslin Institute) and *cytochrome b* sequencing was performed by Chris Palgrave (University of Edinburgh) as part of a Bachelor of Science Honours dissertation (Palgrave 2001). All phylogenetic analyses were carried out by Dr. Simon Goodman (Institute of Zoology, London).

This study is the largest phylogenetic analysis to be undertaken across the Suidae and is currently in preparation for publication (Lowden *et al.* 2004). Until now, the phylogeny of the Javan and Sulawesi warty pigs and the bushpigs (African bushpig and red river hog) has never been reconstructed. Furthermore, adopting this multi-sequence-type approach has enabled key internal nodes within the phylogenetic tree to be resolved where previously single-gene approaches have been incomplete and/or proved inconclusive (Irwin *et al.* 1991; Randi *et al.* 1996; Sulandari *et al.* 1997; Furutani *et al.* 1998; Matthee *et al.* 2001; Palgrave 2001; Randi *et al.* 2002).

Against a background of wide variation in interpretation of the fossil record and discrepancies in existing molecular analyses, the most important conclusion to be

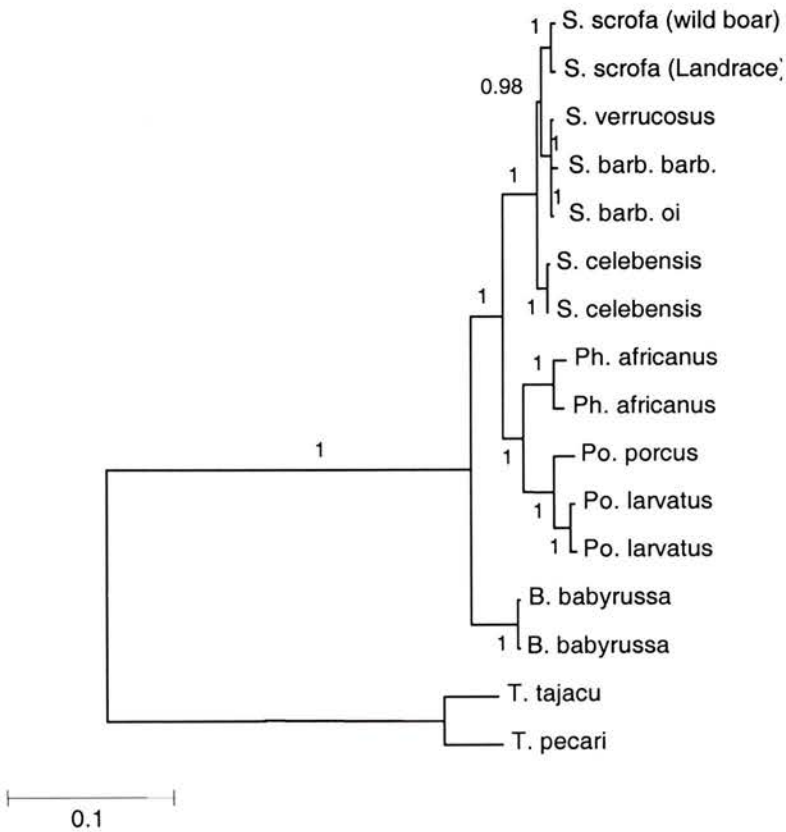
drawn from this analysis is that the Afrotropical suids (bushpigs and warthogs) appear to be monophyletic, having evolved from a common ancestor; as opposed to being the result of two separate evolutionary events (Figure 3.13). It would also appear that the Sulawesi warty pig (*Sus celebensis*) is basal within the *Sus* clade. As expected, babirusa are basal to all other suids. These results compliment and expand on current phylogenetic research into the domestication of pigs (*Sus scrofa*) being undertaken by Greger Larson (University of Oxford) and ongoing work by Professor Chris Moran and colleagues (University of Sydney) to resolve suid phylogeny using *D-loop* sequence (control region) from the mitochondrial genome.

3.8 Comparative Sequencing Overview

The predicted primary protein structure of Light Chain Dynein (LCD), Cyclophilin A (CypA), IκBα and the regulatory domain of NFAT2 was found to be identical between the ASFV-resistant warthog and susceptible domestic pig. Warthog CnAβ has a 3bp and a 30bp insertion relative to domestic pig sequence, equivalent to a one amino acid and a ten amino acid insertion found in two respective splice variants of human CnAβ. In p65 however, three amino acid differences were identified between the domestic pig and warthog sequences. Of particular interest is the presence of a proline residue at position 531 in warthog p65, which is a conserved serine residue in the other pigs and human sequence. This amino acid lies in the transactivation domain 1 (TA1) of p65, which has been well characterised as an important casein kinase II phosphorylation site (ser^{529}) in human p65. This site is involved in secondary level upregulation of NFκB transcriptional activity, distinct from primary IκBα-mediated activation (Bird *et al.* 1997; Wang and Baldwin Jr. 1998; Wang *et al.* 2000; Vermeulen *et al.* 2002). All three identified p65 mutations occur outside the rel homology domain (RHD), which is responsible for interaction with IκBα. Thirty nucleotide differences were identified between the warthog and domestic pig TNFα promoter sequences. In addition, the warthog possesses two single base pair deletions and one six base pair deletion. These nucleotide differences and deletion events determine the existence of a number of putative transcription factor binding domains and transcription factor binding site modules, predicted using components of the

Figure 3.13 – A Molecular Phylogeny of the Suids

Phylogenetic tree illustrating the relationship between the different Genera of Suid, based on sequencing data from the TNFα promoter, *c-Kit* intron seven, a panel of microsatellites and *cytochrome b*. This multi-sequence-type approach has enabled key internal nodes with the phylogenetic tree to be resolved where previously single-gene approaches have proved inconclusive (Randi *et al.* 1996; Palgrave 2001). Importantly, the Afrotropical suids (bushpigs and warthogs) appear to monophyletic, having evolved from a common ancestor; as opposed to being the result of two separate evolutionary events. It appears that the Sulawesi warty pig (*Sus celebensis*) is the basal member of the *Sus* clade. This tree was constructed by Dr. Simon Goodman and is based on Bayesian analysis. Values on nodes represent the probabilities of clades. The scale bar represents percentage sequence divergence (calculated as the sum of the horizontal lines between two species).


S. scrofa (wild boar)
S. scrofa (Landrace)
S. verrucosus
S. barb. barb.
S. barb. oi
S. celebensis
S. celebensis
Ph. africanus
Ph. africanus
Po. porcus
Po. larvatus
Po. larvatus
B. babyrussa
B. babyrussa
T. tajacu
T. pecari
1
0.98
1
1
1
1
1
1
1
1
1
1
1
1
1
0.1

Genomatix software suite. In addition, sequencing of the TNFα promoter was carried out in an additional fourteen animals. These data were combined with that from three other sequencing studies to construct a comprehensive molecular phylogeny of the suids.

Based on these comparative sequencing results, a decision was taken to investigate p65 and the TNFα promoter further; to determine if sequence variation between the warthog and domestic pig is of functional significance.

4 TNFα Promoter Comparative Activity

4.1 Creation of TNFα Promoter-Luciferase Reporter Constructs

Comparative sequencing of the TNFα promoter (Chapter 3) identified thirty nucleotide differences and three deletion events between the domestic pig and warthog sequence. A number of these differences impact on putative transcription factor binding sites and modules predicted using the Genomatix software suite. Therefore, a system was developed to investigate potential functional consequences of these differences. This required coupling the two TNFα promoters to a reporter gene, the output of which could be recorded as a measure of the activity of the promoter in a tissue culture environment.

4.1.1 Luciferase Reporters

In order to measure activity of the TNFα promoter *in vitro*, a robust and easily quantifiable reporter gene was required. Luciferase was selected as it can be readily measured from a simple cell lysis preparation (Section 2.4.4). Luciferase concentration is quantified by adding its substrate and measuring the luminescent signal produced as a result of enzymatic properties. There are two forms of the luciferase gene commonly available; one is derived from the firefly (*Photinus pyralis*) and the other from the sea pansy (*Renilla reniformis*). Firefly and renilla luciferases are evolutionarily distinct and as such have very different structures and substrates (Zhuang *et al.* 2001). They can therefore be assayed by separate reactions (Figure 4.1). In the following experiments, firefly luciferase was used as the experimental reporter and renilla luciferase was used as a control reporter.

4.1.2 Introduction of Restriction Sites by PCR

First-round TNFα promoter PCR products from the sequencing phase of the work (Chapter 3) were diluted to 1:500 in water with 1µl then used as a template in a 25µl PCR reaction containing 1xPCR buffer, 2mM dNTPs, 2mM $MgCl_2$, 0.7U *High*

Figure 4.1 – Firefly and *Renilla* luciferase reactions

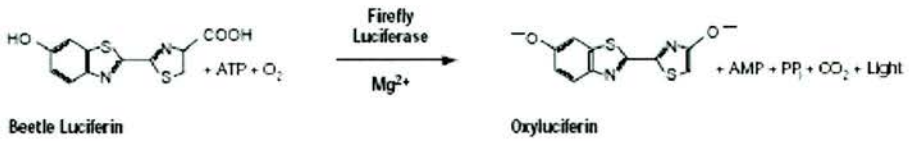
Two forms of the luciferase gene are commonly available, derived from the firefly (*Photinus pyralis*) and the sea pansy (*Renilla reniformis*). Firefly and *Renilla* luciferases are evolutionarily distinct and as such have different structures and substrates, however both reactions emit a similar luminescent signal which can be measured.

A Firefly luciferase, a 61kDa monomeric protein, catalyses the oxidation of beetle luciferin into oxyluciferin, a process which requires ATP, Mg^{2+} and oxygen.

B Coelenterazine is converted to coelenteramide by *Renilla* luciferase, a 36kDa monomeric protein, a process which requires oxygen.

Figures adapted from Promega Corporation Technical Manual No. 040
'Dual-Luciferase Reporter Assay System'

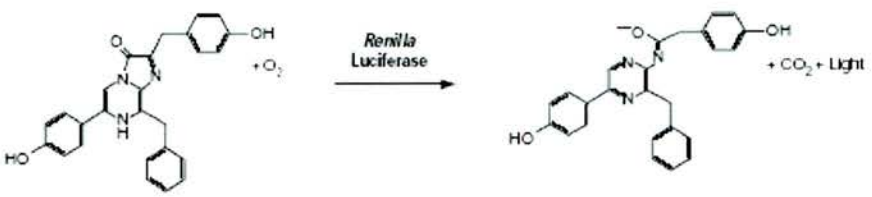
A



Beetle Luciferin

Oxyluciferin

B



Coelenterazine

Coelenteramide

Fidelity DNA polymerase (Roche Diagnostics) and 20pmol *TNFα* restriction primers to introduce specific restriction sites (underlined): Forward *Xho* I (5'-TAA AGC TCG AGT CTC TGA AAT GCT-3') and reverse *Hin*d III (5'-TGT AAG CTT CAG AGG GGC TCA G-3'). Each ~1600bp PCR product was run on a 1% agarose gel, excised with a scalpel blade and gel-extracted as previously described.

4.1.3 Creation of $P_{TNF\alpha}$-Luciferase Single-Reporter Plasmids

To create single reporter plasmids, the pGL3-Basic vector (Promega) was used as a backbone. This contains the gene encoding firefly luciferase (FL), downstream from a multiple cloning site into which a promoter can be inserted.

5μg pGL3-Basic vector was digested using *Xho* I and *Hin*d III in SuRE/Cut Buffer B and the open ends dephosphorylated as previously described (see section 2.3.3). $P_{TNF\alpha}$ PCR products with PCR-introduced restriction sites were also digested with *Xho* I and *Hin*d III in SuRE/Cut Buffer B. The promoters were ligated into the open pGL3-Basic vector to produce single reporter plasmids containing either the domestic pig or warthog $P_{TNF\alpha}$ driving firefly luciferase production ($P_{TNF\alpha}$-FL) (Figure 4.2a).

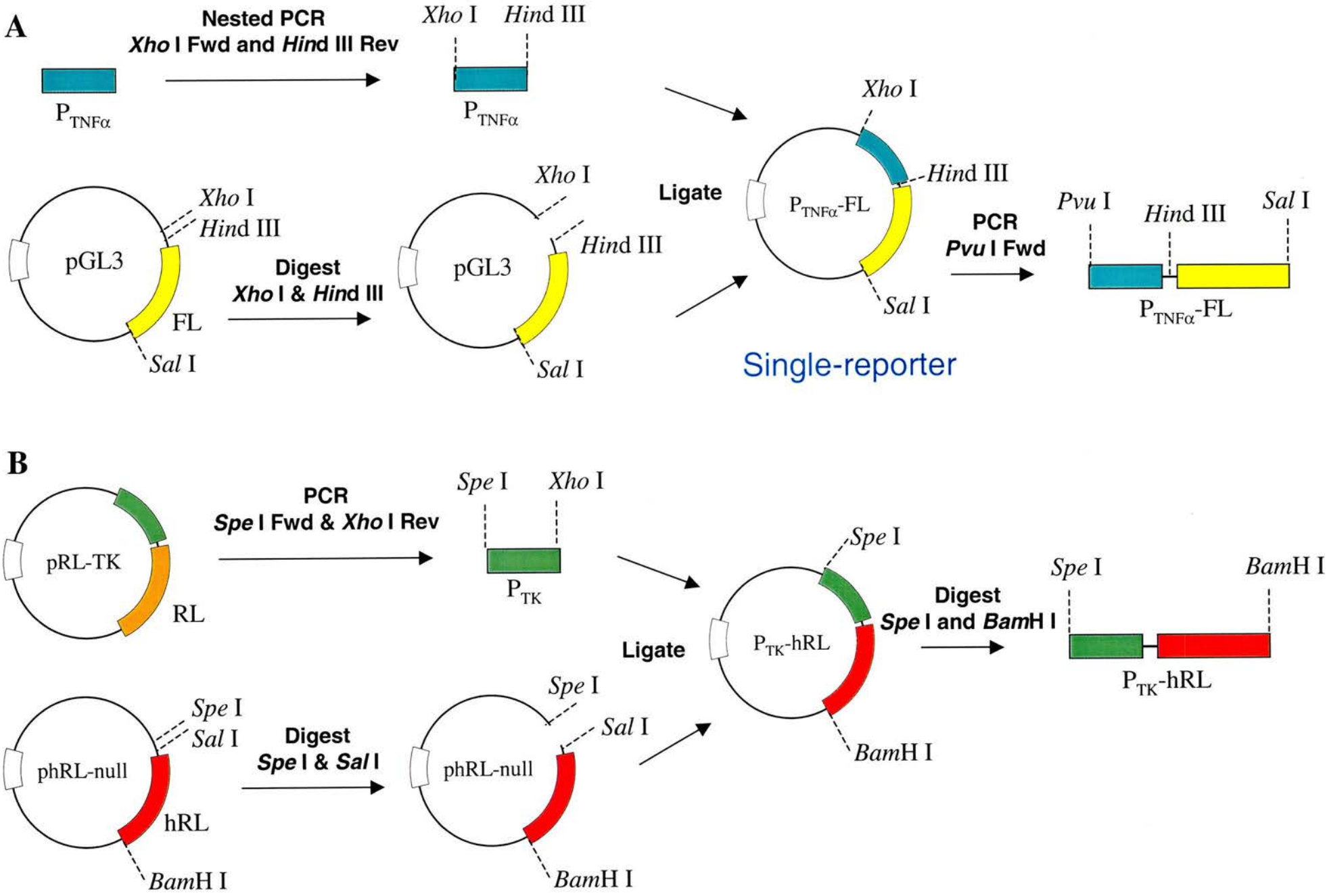
4.1.4 Transient Transfection of Single-Reporter Plasmids

In order to verify that the single-reporter cassettes were working correctly, the $P_{TNF\alpha}$-luciferase plasmids were transfected into murine Hepa 1 cells (Section 2.4.1.1) and harvested in lysis buffer 48 hours post-transfection. The lysate was assayed to check for luciferase activity and produced a strong luminescent signal.

Figure 4.2 – Cloning Strategy to Develop Single-Reporter Cassettes

A A nested PCR was performed on the warthog or domestic pig TNFα promoter PCR product from the comparative sequencing project to introduce *Xho* I and *Hin*d III restriction sites. The TNFα promoter was digested with *Xho* I and *Hin*d III as was the pGL3 plasmid. The two were ligated together to form the TNFα promoter–firefly luciferase ($P_{TNF\alpha}$-FL) single reporter plasmids. The entire $P_{TNF\alpha}$-FL cassette was then amplified from the plasmid using a forward primer which introduced a *Pvu* I site, for insertion into the dual-reporter plasmid.

B The thymidine kinase promoter (P_{TK}) was amplified from the pRL-TK plasmid using primers which introduced a *Spe* I and *Xho* I restriction site. The P_{TK} was digested with *Spe* I and *Xho* I and the phRL-null plasmid with *Spe* I and *Sal* I. The two were ligated together to form a P_{TK}-humanised renilla luciferase (P_{TK}-hRL) reporter construct (*Sal* I and *Xho* I sites were destroyed in the process). The P_{TK}-hRL cassette was then removed from the plasmid by digestion with *Spe* I and *Bam*H I, ready for insertion into the dual-reporter plasmid.


A
Nested PCR
Xho I Fwd and Hind III Rev
Xho I
Hind III
P TNFα
P TNFα
Xho I
Hind III
pGL3
FL
Sal I
Digest
Xho I & Hind III
Xho I
Hind III
pGL3
Sal I
Ligate
Xho I
Hind III
P TNFα-FL
Sal I
PCR
Pvu I Fwd
Pvu I
Hind III
Sal I
P TNFα-FL
Single-reporter
B
pRL-TK
RL
PCR
Spe I Fwd & Xho I Rev
Spe I
Xho I
P TK
phRL-null
hRL
BamH I
Spe I
Sal I
Digest
Spe I & Sal I
Spe I
Sal I
phRL-null
BamH I
Ligate
Spe I
P TK-hRL
BamH I
Digest
Spe I and BamH I
Spe I
BamH I
P TK-hRL

4.2 Creation of Dual-Reporter Plasmids

When using a single-luciferase reporter, results are usually normalised by dividing luciferase activity by the total protein concentration to account for variation in cell number. However, this method can lead to inaccuracy, due to variation in plasmid copy number within cells containing the different experimental constructs. Furthermore, it was considered that the difference in activity between our experimental promoters may be subtle and require a system offering higher resolution. It was therefore decided to create a dual-reporter plasmid which contained both experimental and control luciferase reporter cassettes. Under these conditions, a ratio of experimental luciferase (firefly) to control luciferase (renilla) would give a 'self-normalising' 'per-plasmid' value, regardless of any variation in cell or plasmid number. It also was anticipated that use of this system would facilitate detection of subtle variation between experimental constructs.

As no suitable commercial vector was available, novel dual-luciferase reporter constructs were designed and created. These dual-reporters contained the $P_{TNF\alpha}$-firefly luciferase expression cassette as well as a control promoter-renilla luciferase cassette. In order to create these constructs, the pBudCE4.1 (Invitrogen) plasmid was used as a backbone, as it contains two multiple cloning sites. This plasmid encodes the gene for zeocin resistance, rather than that for ampicillin or G418 resistance. Figures 4.2 and 4.3 give an overview of the cloning strategy detailed below.

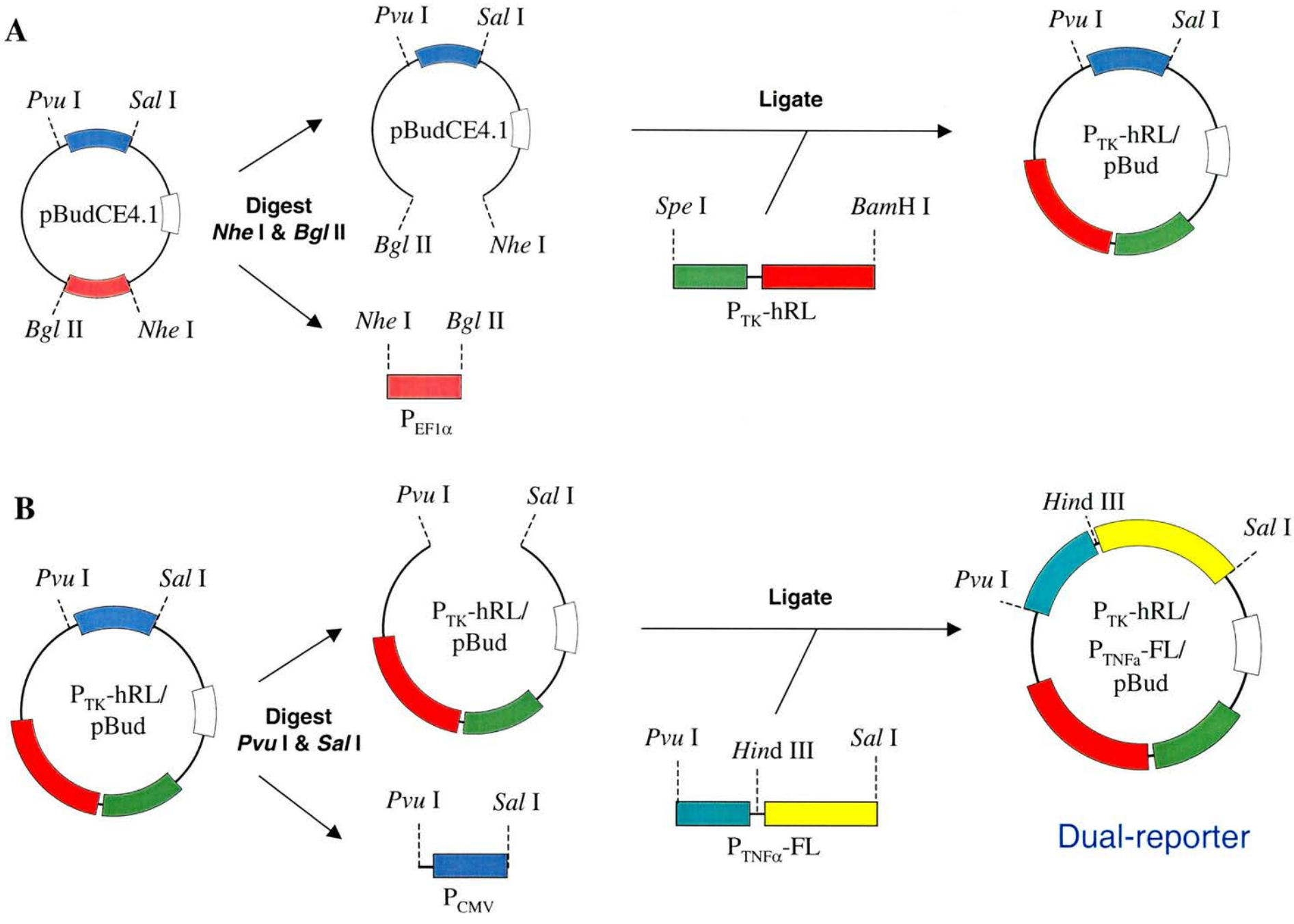
4.2.1 Experimental $P_{TNF\alpha}$-Firefly Luciferase Cassette

The domestic pig and warthog $P_{TNF\alpha}$-FL cassettes were amplified from their respective pGL3-Basic single-reporter plasmids using 10ng DNA as a template in a 25µl PCR reaction containing 1xPCR buffer, 2mM dNTPs, 2mM $MgCl_2$, 0.7U *High Fidelity* DNA polymerase and 20pmol *Pvu* I Fwd primer (5'-AGT CGA TCG GCC AGA ACA TT-3') and Rev primer (5'-TGA CTG GGT TGA AGG CTC TC-3'). This PCR product (~3.5kbp) was resolved on an agarose gel, gel-extracted and

Figure 4.3 – Cloning Strategy to Create the Dual-reporter Plasmids

A The pBudCE4.1 plasmid was digested with *Nhe* I and *Bgl* II to remove the $P_{EF1\alpha}$ already present. The previously created P_{TK}-hRL cassette (see Figure 29b) was ligated into the open plasmid to create an intermediate P_{TK}-hRL/pBud construct. During the process despite having complimentary overhangs, all four restriction sites involved were destroyed (*Nhe* I, *Spe* I, *Bgl* II, and *Bam*H I).

B The intermediate P_{TK}-hRL/pBud construct was digested with *Pvu* I and *Sal* I to remove the P_{CMV} already present. The $P_{TNF\alpha}$-FL cassettes amplified from the single-reporter plasmids (see Figure 29a) were also digested with *Pvu* I and *Sal* I. These were ligated into the open P_{TK}-hRL/pBud construct to form two final P_{TK}-hRL/$P_{TNF\alpha}$-FL/pBud plasmids, one containing the domestic pig TNFα promoter and the other the warthog TNFα promoter.


A
Pvu I
Sal I
pBudCE4.1
Bgl II
Nhe I
Digest
Nhe I & Bgl II
Pvu I
Sal I
pBudCE4.1
Bgl II
Nhe I
Nhe I
Bgl II
P EF1α
Ligate
Spe I
BamH I
P TK-hRL
Pvu I
Sal I
P TK-hRL/
pBud
B
Pvu I
Sal I
P TK-hRL/
pBud
Digest
Pvu I & Sal I
Pvu I
Sal I
P TK-hRL/
pBud
Pvu I
Sal I
P CMV
Ligate
Pvu I
Hind III
Sal I
P TNFα-FL
Hind III
Sal I
Pvu I
P TK-hRL/
P TNFa-FL/
pBud
Dual-reporter

digested using *Pvu* I and *Sal* I (the latter site was already present in the sequence) in SuRE/Cut Buffer H to prepare the cassette for subsequent use (Figure 4.2a).

4.2.2 Control Promoter-Reporter Cassette

The herpes simplex virus (HSV) thymidine kinase promoter (P_{TK}) was selected as the control promoter because it is a relatively small (~760bp) constitutive promoter, which is not as highly active as the P_{CMV} and P_{SV40} promoters which are commonly used in expression constructs. It is also relatively uninfluenced by other cellular processes and is not as susceptible to a gradual decrease in activity/decay over time (unlike P_{CMV}) (Promega technical services, personal communication).

The P_{TK} was amplified by PCR from the pRL-TK vector (Promega), where it was driving expression of the native renilla luciferase gene. The PCR reaction used 10ng pRL-TK vector as template in a 25µl PCR reaction containing 1xPCR buffer, 2mM dNTPs, 2mM $MgCl_2$, 0.7U *High Fidelity* DNA polymerase and 20pmol *Spe* I Fwd primer (5'-TGG ACT AGT GCT CAC ATG GC-3') and *Xho* I Rev primer (5'-AGT GCC TCG AGA CCA ACT TC-3'). The resulting PCR product was run out on an agarose gel and extracted. The purified PCR product was digested using *Spe* I and *Xho* I in SuRE/Cut Buffer H to prepare the P_{TK} for insertion into the next vector.

The above restriction sites enabled the P_{TK} to be inserted into the multiple cloning site of phRL-null vector (Promega), which contains a 'humanised' renilla luciferase (hRL), a synthetic gene lacking ~300 of the consensus sequence transcription factor binding sites found naturally in the coding sequence of the native gene. This prevents anomalous transcription behaviour and allows for more efficient activity of the control promoter-reporter system (Zhuang *et al.* 2001). The phRL-null vector was digested using *Spe* I and *Sal* I in SuRE/Cut Buffer H, the open ends dephosphorylated and the P_{TK} ligated into it. During the ligation reaction, the complementary *Xho* I overhang on P_{TK} and the *Sal* I overhang on phRL-null fused together, destroying both restriction sites in the process.

In order to remove the entire P_{TK}-hRL cassette, ready for ligation into the pBudCE4.1 dual reporter backbone, a *Spe* I and *Bam*H I digestion was performed with SuRE/Cut Buffer M to release a product ~2.2kbp in length (Figure 4.2b).

4.2.3 Creating the Dual-Reporter Plasmid

In order to receive the P_{TK}-hRL cassette, 5µg pBudCE4.1 vector (subsequently referred to as pBud) was digested using *Nhe* I and *Bgl* II in SuRE/Cut Buffer M. This removed the $EF_{1\alpha}$ promoter already present in the vector. This digest was run on an agarose gel and the band corresponding to the vector without $P_{EF1\alpha}$ (~3.4kb) was excised, gel-extracted and the open ends dephosphorylated. The P_{TK}-hRL cassette was ligated into the open pBud vector to produce the P_{TK}-hRL/pBud plasmid. This process involved both the complimentary *Spe* I and *Nhe* I and the complimentary *Bam*H I and *Bgl* II overhangs ligating together. In the process all four restriction sites were destroyed (Figure 4.3a).

To insert the experimental $P_{TNF\alpha}$-FL cassette, 5µg of the P_{TK}-hRL/pBud plasmid was digested with *Pvu* I and *Sal* I in SuRE/Cut Buffer H. This removed the CMV promoter (P_{CMV}) already present in the plasmid. The digestion was run out on an agarose gel and the band corresponding to the plasmid without the P_{CMV} (~5kb) excised, gel extracted and the open ends dephosphorylated. The $P_{TNF\alpha}$-FL cassettes were then ligation-inserted to create two final plasmids, one containing the domestic pig $P_{TNF\alpha}$ and the other containing the warthog $P_{TNF\alpha}$ (Figure 4.3b). Sequencing was carried out to confirm the identity of the plasmids.

4.2.4 Transfection of Cells with Dual-Reporter Constructs

The two dual-reporter plasmids containing either the domestic pig or warthog $P_{TNF\alpha}$-FL cassette together with the control P_{TK}-hRL cassette were transfected into Hepa 1 cells as described in Chapter 2. These transfections were carried out in duplicate. One flask of each plasmid type was harvested 48 hours post-transfection and assayed

for transient luciferase activity to ensure that both cassettes were functioning correctly. High levels of both firefly and renilla luciferase were detected in both samples. The second flask of each plasmid type was placed under selection using 200μg/ml Zeocin over a ten-day period to obtain stably transfected colonies. These cells were then pooled, cultured to increase cell number and frozen down for future stocks.

4.3 Experimental Design

Stably transfected cells were cultured in twelve-well tissue culture plates to achieve ~90% confluence on the day of the experiment. In order to assess differences in the TNFα promoter activity, it was first necessary to induce the cells to produce a cytokine response. Three inducing agents with different modes of action were selected to elicit an immune-type response from the cells. These agents (which have previously been successfully used for this purpose) were lipopolysaccharide (LPS), tumour necrosis factor alpha (TNFα) and phorbol-12-myristate-13-acetate (PMA) (de Martin *et al.* 1993; Powell *et al.* 1996; Abu-Amer *et al.* 1997; Haudek *et al.* 1998; Revilla *et al.* 1998; Wang and Baldwin Jr. 1998; Tait *et al.* 2000; Wang *et al.* 2000). The mode of action of the above agents is described below in Section 4.4. All inducing agents were diluted in tissue culture medium and applied to the cells in culture for five hours. After induction, the medium was removed, the cells washed with PBS and total protein harvested in passive lysis buffer (Promega). Firefly and renilla luciferase activity was measured using the Dual-Reporter Assay System (Promega), as described in Section 2.4.4.

All treatments were carried out in duplicate wells and the subsequent assays were carried out in triplicate. Once the average results for each well had been calculated, the results of the duplicate wells were calculated as an average of the pair (n=2). In addition to the transfected cells, untransfected Hepa cells were also harvested to detect any background reading from the luminometer.

4.4 Inducing Agents and Dose Range

In order to elicit an immune-type response from the cell, LPS was used at 1, 5, 10, 20, 30, 50, 100μg/ml whereas TNFα and PMA were used at 0.1, 1, 5, 10, 30, 50 and 100ng/ml. These ranges of concentrations were selected as representative of those previously used in the studies referenced above in Section 4.3 and included doses sufficient to generate a maximum response.

4.4.1 Lipopolysaccharide (LPS)

Lipopolysaccharide (LPS) is a glycolipid component from the cell wall of Gram-negative bacteria and a potent immunogenic antigen, acting primarily through the CD14-Toll-like receptor (TLR)4 complex to activate NFκB (Bäckhed and Hornef 2003), although CD14-TLR4-mediated NFκB-independent mechanisms have recently been described (Fan *et al.* 2004). The preparation used in this study was derived from *Escherichia coli* 0127:B8 (Sigma Aldrich) and is the same as that used in a previously mentioned ASFV *in vitro* cell challenge by Powell *et al.* (1996). Furthermore, LPS was used to assess differences between human and baboon TNFα promoter activity in response to induction (Haudek *et al.* 1998).

4.4.2 Tumor Necrosis Factor Alpha (TNFα)

Recombinant human TNFα (Sigma Aldrich) was used in this study, which acts on both murine and human cell lines (Sigma technical services). The soluble TNFα used here primarily induced a TNFR1 response, as most TNFR2 activity is attributed to binding membrane-bound TNFα (MacEwan 2002).

4.4.3 Phorbol-12-myristate-13-acetate (PMA)

Phorbol-12-myristate-13-acetate (PMA) (Sigma Aldrich) is a phorbol ester analogous to diacylglycerol (DAG). DAG is usually produced by the signal-induced

hydrolysis of membrane lipid phosphatidylinositol (PIP2) into the second messengers, DAG and inositol triphosphate (IP_3). DAG is a direct allosteric activator of protein kinase C (PKC). IP_3 causes the release and mobilisation of intracellular Ca^{2+} and activation of numerous signalling pathways, including NFAT and indirect PKC activation (Basta *et al.* 2001; Schaefer 2004).

4.5 Dual-Reporter Induction Results

During these experiments, it became clear that the activity of the domestic pig TNFα promoter ($P_{TNF\alpha}$) was consistently and significantly higher than that of the warthog ($p<0.01$). This was the case when the cells were both uninduced and induced. The raw data can be seen in Table 4.1. Professional assistance with statistical analyses described below was kindly given by Anthea Spingbett (Roslin Institute). Data from each experiment is described in detail below.

4.5.1 Control Renilla Luciferase Activity

The P_{TK}-renilla luciferase cassette was incorporated into the dual-reporter plasmids as a constitutively-expressed control reporter. As such, its activity was not expected to change during the course of the experiment. Renilla luciferase activity from all experiments were plotted against the concentration of inducing agent (LPS, μg/ml; TNFα and PMA, ng/ml) (Figure 4.4). Apart from two extremely low values for both replicates of the same sample (warthog, 100ng/ml PMA), renilla luciferase levels did not vary significantly and were subsequently used to normalise the experimental firefly luciferase data. The two outlying renilla luciferase results could not be explained and were omitted from subsequent analyses. Any minor fluctuation in renilla luciferase, as a result of the induction, can be ignored as its effect is assumed to be identical for both experimental constructs.

Figure 4.4 – Control Renilla Luciferase Activity

Renilla luciferase levels plotted against the concentration of inducing agents (LPS, μg/ml; TNFα and PMA, ng/ml) for dual-reporter constructs containing either the domestic pig (SS) or warthog (PA) TNFα promoter. Apart from two extremely low values for both replicates of one sample (warthog, 100ng/ml PMA), renilla levels did not vary significantly during the course of the experiment. These outliers cannot be easily explained and were omitted from subsequent analyses.

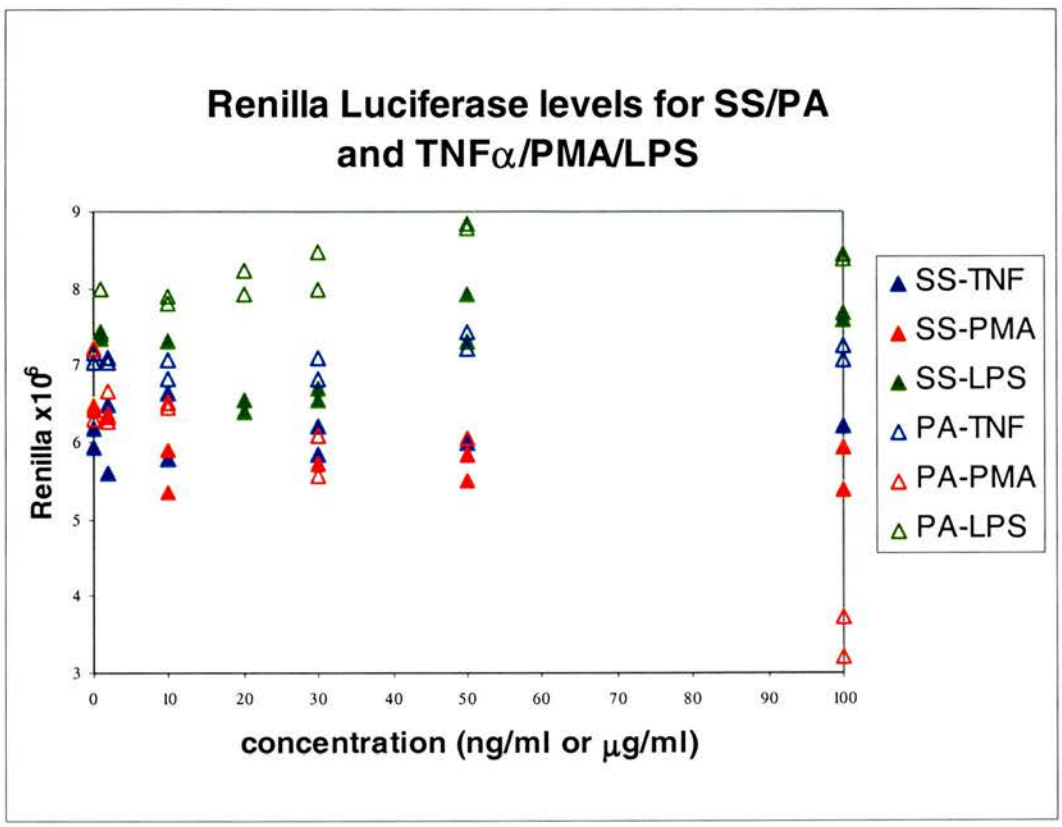

Renilla Luciferase levels for SS/PA and TNFα/PMA/LPS
Renilla x10^6
9
8
7
6
5
4
3
0
10
20
30
40
50
60
70
80
90
100
concentration (ng/ml or μg/ml)
SS-TNF
SS-PMA
SS-LPS
PA-TNF
PA-PMA
PA-LPS

Table 4.1

Spreadsheet of the data obtained from the *in vitro* comparison of the domestic pig and warthog TNFα promoters (within dual-luciferase reporter constructs). Results were normalised by dividing experimental firefly luciferase activity by control renilla luciferase. Each condition was carried out in duplicate and each assay performed in triplicate.

Sample	Species	Treatment(s)	Concn	Firefly 1	Firefly 2	Firefly 3	mean Firely	Renilla 1	Renilla 2	Renilla 3	Mean Renilla	mean firefly/mean renilla (x1000)
1	SS	TNF	0.1	101573	87913	99699	96395	6508503	5838617	6219628	6188916	15.58
2	SS	TNF	0.1	77417	81844	71257	76839	6323716	5860240	5650387	5944781	12.93
3	SS	TNF	2	113076	116087	103887	111017	7006153	6241508	6193432	6480364	17.13
4	SS	TNF	2	115019	105041	95516	105192	6109934	5560255	5127710	5599300	18.79
5	SS	TNF	10	174019	161782	141899	159233	7284266	6253791	6346386	6628148	24.02
6	SS	TNF	10	119128	110346	113179	114218	6239178	5598104	5531180	5789487	19.73
7	SS	TNF	30	148981	120704	121437	130374	6177208	5696307	5623299	5832271	22.35
8	SS	TNF	30	149640	133440	124280	135787	6731726	6138085	5807131	6225647	21.81
9	SS	TNF	50	146025	123047	121361	130144	6364931	5924741	5918262	6069311	21.44
10	SS	TNF	50	135029	130896	116398	127441	6358678	5869575	5755305	5994519	21.26
11	SS	TNF	100	139801	131234	118255	129763	6765371	5868123	6044873	6226122	20.84
12	SS	TNF	100	155759	136311	130436	140835	6632277	5741125	5418717	5930706	23.75
13	SS	PMA	0.1	123618	107789	107323	112910	6968980	6297535	6224673	6497063	17.38
14	SS	PMA	0.1	128888	114811	115915	119871	6931638	6147327	6179437	6419467	18.67
15	SS	PMA	2	132266	111410	117398	120358	6963190	6058539	6032955	6351561	18.95
16	SS	PMA	2	127552	115641	111175	118123	6985156	5952667	6177796	6371873	18.54
17	SS	PMA	10	113960	109268	102249	108492	6409296	5675407	5610306	5898336	18.39
18	SS	PMA	10	111213	101476	96327	103005	5800880	5137874	5172622	5370459	19.18
19	SS	PMA	30	120828	103443	94729	106333	6418333	5626741	5109475	5718183	18.60
20	SS	PMA	30	123575	109861	102902	112113	6255615	5442877	5459530	5719341	19.60
21	SS	PMA	50	114784	108469	112681	111978	6038349	5715480	5745437	5833089	19.20
22	SS	PMA	50	118087	110619	109370	112692	5809730	5297664	5421493	5509629	20.45
23	SS	PMA	100	117896	102845	107009	109250	5927088	5131624	5146097	5401603	20.23
24	SS	PMA	100	137587	116560	106686	120278	6616829	5735772	5501013	5951205	20.21
25	SS	LPS	1	198395	161387	159836	173206	8149019	6790305	7077840	7339055	23.60
26	SS	LPS	1	192061	165414	167647	175041	8252924	6944325	7005946	7401065	23.65
27	SS	LPS	10	206943	177890	179633	188155	8159197	7197828	6600224	7319083	25.71
28	SS	LPS	10	253512	223821	218228	231854	7925526	7144098	6887024	7318883	31.68
29	SS	LPS	20	196719	181715	181875	186770	6361206	6410017	6431624	6400949	29.18
30	SS	LPS	20	185093	173698	158130	172307	7148066	6389161	6093456	6543561	26.33
31	SS	LPS	30	195811	181485	182246	186514	6992262	6425374	6705647	6707761	27.81
32	SS	LPS	30	208906	183381	179006	190431	7111052	6250105	6292111	6551089	29.07
33	SS	LPS	50	255447	253491	237920	248953	8381129	7729501	7709566	7940065	31.35
34	SS	LPS	50	203851	212575	210814	209080	7362173	7254620	7331142	7315978	28.58
35	SS	LPS	100	266282	234297	232992	244524	8194263	7364844	7459753	7672953	31.87
36	SS	LPS	100	280656	250251	225221	252043	8426098	7060175	7332479	7606251	33.14
37	SS	unstimulated		137901	123214	128887	130001	7465588	6423120	6227684	6705464	19.39
38	SS	unstimulated		123807	108906	108341	113685	6491562	5710287	5808485	6003445	18.94
39	PA	TNF	0.1	76380	59825	65649	67285	7764536	6680320	6723286	7056047	9.54
40	PA	TNF	0.1	78815	67476	71559	72617	7960616	6863770	6647152	7157179	10.15
41	PA	TNF	2	95247	82622	79971	85947	7652658	6661453	6801951	7038687	12.21
42	PA	TNF	2	90933	87201	87315	88483	7478210	7138617	6678972	7098600	12.46
43	PA	TNF	10	99375	88490	91366	93077	7682299	6928738	6605484	7072174	13.16
44	PA	TNF	10	108527	93746	90058	97444	7405506	6563409	6467085	6812000	14.30
45	PA	TNF	30	101717	78169	86282	88723	7442392	6459552	6532807	6811584	13.03
46	PA	TNF	30	112449	92066	91205	98573	7625397	6862727	6854882	7114335	13.86
47	PA	TNF	50	117095	102149	102482	107242	8195664	7049532	7029079	7424758	14.44
48	PA	TNF	50	110110	98206	98793	102370	7535491	6954866	7194575	7228311	14.16
49	PA	TNF	100	97365	93618	92340	94441	7510647	6830525	6875055	7072076	13.35
50	PA	TNF	100	111759	96253	95624	101212	8017518	6875537	6848621	7247225	13.97
51	PA	PMA	0.1	88865	75898	77337	80700	7997077	6712402	6988135	7232538	11.16
52	PA	PMA	0.1	74755	68858	65385	69666	6524402	6362793	6063824	6317006	11.03
53	PA	PMA	2	75318	69481	74325	73041	7130967	6394492	6444928	6659796	10.97
54	PA	PMA	2	84713	75514	72961	77729	6829124	5943867	6045334	6272775	12.39
55	PA	PMA	10	83967	70693	74247	76302	6989129	6181200	6226509	6465613	11.80
56	PA	PMA	10	88726	71216	66151	75364	6995702	6228591	6297263	6507185	11.58
57	PA	PMA	30	79592	69032	70546	73057	6734122	5689310	5886120	6103184	11.97
58	PA	PMA	30	74658	61684	62828	66390	5968086	5301872	5437146	5569035	11.92
59	PA	PMA	50	76337	64957	66256	69183	6651531	5870447	5644041	6055340	11.43
60	PA	PMA	50	65979	58477	57811	60756	6381752	5414381	4769600	5521911	11.00
61	PA	PMA	100	52797	54132	56462	54464	3706654	2991697	2903020	3200457	17.02
62	PA	PMA	100	69963	64352	64280	66198	3737576	3368290	4091998	3732621	17.74
63	PA	LPS	1	124027	122384	112493	119635	7420499	7412125	7520802	7451142	16.06
64	PA	LPS	1	136516	118816	120329	125220	8713966	7860036	7400926	7991643	15.67
65	PA	LPS	10	141970	124419	116340	127576	8480514	7345079	7608187	7811260	16.33
66	PA	LPS	10	142612	120596	124804	129337	8724364	7483136	7500491	7902664	16.37
67	PA	LPS	20	158281	129470	126737	138163	8429578	7774435	7587431	7930481	17.42
68	PA	LPS	20	161419	131957	129708	141028	9144454	7886152	7672850	8234485	17.13
69	PA	LPS	30	151313	123690	128732	134578	8753494	7731752	7490611	7991952	16.84
70	PA	LPS	30	166767	133716	145241	148575	9407604	7677390	8358213	8481069	17.52
71	PA	LPS	50	164072	148577	154439	155696	9086392	8622478	8649866	8786245	17.72
72	PA	LPS	50	175550	151690	143164	156801	9800716	8504980	8236976	8847557	17.72
73	PA	LPS	100	176466	149440	143166	156357	9142630	8219912	7964415	8442319	18.52
74	PA	LPS	100	157909	151284	147173	152122	9137720	8042933	7942131	8374261	18.17
75	PA	unstimulated		88710	74844	76628	80061	6945952	6231044	6201648	6459548	12.39
76	PA	unstimulated		98147	81707	77367	85740	7621789	6347773	6439989	6803184	12.60

4.5.2 Inherent Difference in Activity

In all samples, regardless of induction, the domestic pig TNFα promoter was significantly more active than the warthog TNFα promoter ($p<0.01$) (Figure 4.5).

4.5.3 LPS Treatment of Transfected Cells

When the cells were treated with LPS, the domestic pig $P_{TNF\alpha}$ was significantly more responsive to induction (seen as a steeper induction curve) than the warthog $P_{TNF\alpha}$ ($p<0.05$) (Figure 4.5a).

The best fitting model for these data (firefly luciferase normalised to renilla luciferase) is two straight lines of differing slopes and intercepts for the domestic pig and warthog $P_{TNF\alpha}$. Where x = dose and y = response:

Domestic pig $P_{TNF\alpha}$ response to LPS is: $y = 26.0 + 0.07x$

Warthog $P_{TNF\alpha}$ response to LPS is: $y = 16.3 + 0.02x$

The difference in gradient between the slopes is significant ($p<0.05$) as is the difference between intercepts ($p<0.01$; the basal level of expression). Across the range of concentrations tested here, the response of the domestic pig $P_{TNF\alpha}$ was always significantly higher than that of the warthog.

4.5.4 TNFα Treatment of Transfected Cells

When TNFα was used as an inducing agent, both the domestic pig and warthog $P_{TNF\alpha}$ responded significantly ($p<0.05$). Both promoters responded in a similar manner to induction (Figure 4.5b), although they started at significantly different basal levels ($p<0.01$).

Figure 4.5 – Response of Dual-Luciferase Reporters to Induction

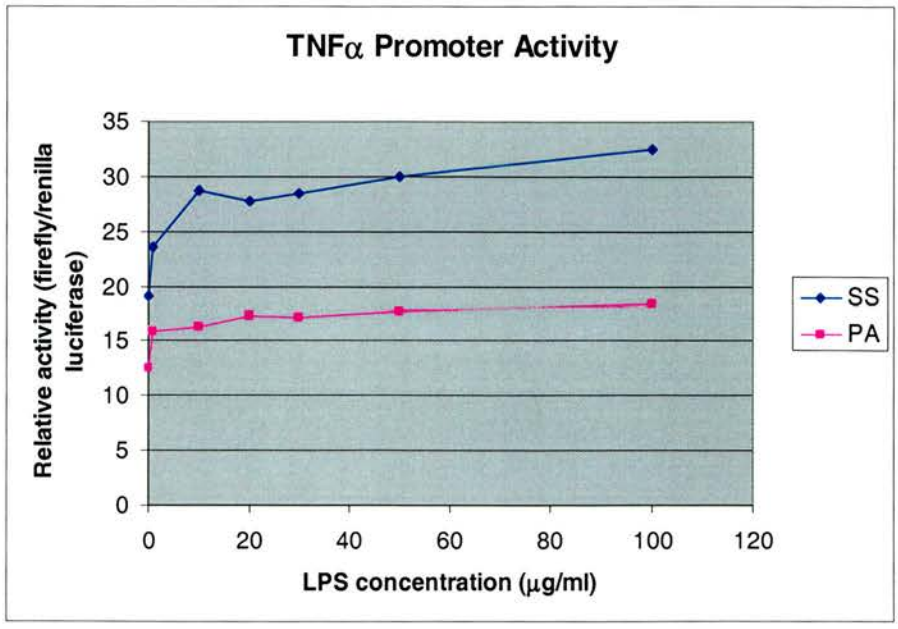
Experimental firefly luciferase results were normalised to control renilla luciferase and the data used to plot the graphs described below.

A The activity of domestic pig and warthog TNFα promoters in response to increasing concentrations of LPS. These results indicate that the domestic pig TNFα promoter responds significantly more strongly than the warthog ($p<0.01$) to LPS induction. This is seen as a steeper induction gradient.

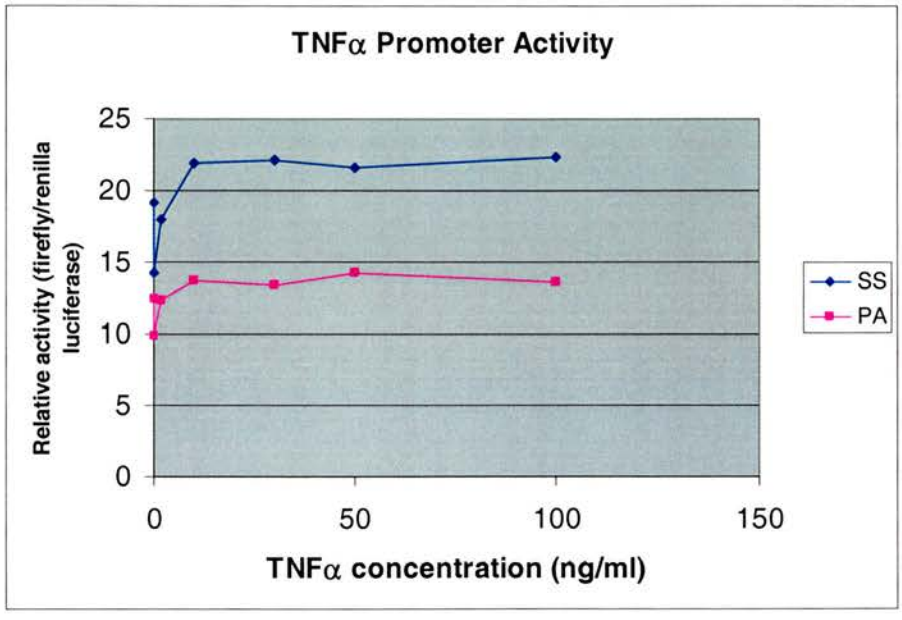
B Both domestic pig and warthog TNFα promoters respond in a similar manner and magnitude to TNFα induction. However they start at significantly different levels ($p<0.01$).

C Neither the domestic pig or warthog TNFα promoter responded significantly to induction with PMA. However, they maintained their significantly different intrinsic levels of activity throughout the experiment ($p<0.01$).

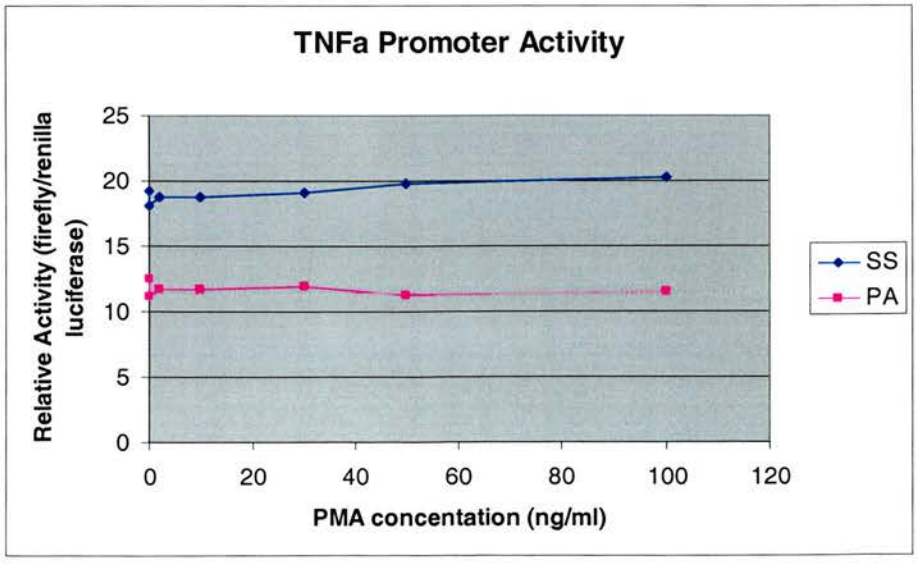
A


TNFα Promoter Activity
Relative activity (firefly/renilla luciferase)
35
30
25
20
15
10
5
0
0
20
40
60
80
100
120
LPS concentration (μg/ml)
SS
PA

B


TNFα Promoter Activity
Relative activity (firefly/renilla luciferase)
25
20
15
10
5
0
0
50
100
150
TNFα concentration (ng/ml)
SS
PA

C


TNFa Promoter Activity
Relative Activity (firefly/renilla luciferase)
25
20
15
10
5
0
0
20
40
60
80
100
120
PMA concentation (ng/ml)
SS
PA

The best fitting model for these data (firefly luciferase normalised to renilla luciferase) has separate intercepts for the two species, which are significantly different ($p<0.01$), but share a common slope and quadratic term. The regression equations are as follows, where x = dose and y = response:

Domestic pig $P_{TNF\alpha}$ response to TNFα is: $y = 17.8 + 0.146x - 0.0011x^2$

Warthog $P_{TNF\alpha}$ response to TNFα is: $y = 10.7 + 0.146x - 0.0011x^2$

The linear and quadratic terms are both significantly different from zero ($p<0.05$; i.e. TNFα induces a significant response) and are the same in both species. As the domestic pig $P_{TNF\alpha}$ has a significantly higher intercept, so the responses are uniformly higher than those of the warthog $P_{TNF\alpha}$ over the entire range. Interestingly, at the lowest TNFα dose (0.1ng/ml), the response for both species drops below that seen in unstimulated cells.

4.5.5 PMA Treatment of Transfected Cells

During the experiment using PMA as the inducing agent, neither the domestic pig nor warthog $P_{TNF\alpha}$ responded significantly to induction; this can be seen in Figure 4.5c.

The best fitting model for these data (firefly luciferase normalised to renilla luciferase) is two straight lines of differing intercepts for the domestic pig and warthog $P_{TNF\alpha}$. Where x = dose and y = response:

Domestic pig $P_{TNF\alpha}$ response to PMA is: $y = 18.4 + 0.028x$

Warthog $P_{TNF\alpha}$ response to PMA is: $y = 11.5 + 0.00005x$

The difference between intercepts (basal expression) is significant ($p<0.01$) and neither construct varied from this throughout the course of the induction. Possible reasons for this lack of induction will be discussed later.

4.6 Summary – Comparative TNFα Promoter Activity

Within this experimental system, it appears that sequence differences identified between the domestic pig and warthog TNFα promoters (Chapter 3) have a significant impact on function *in vitro*.

The domestic pig TNFα promoter appears to be significantly more active than that of the warthog, even without induction ($p<0.01$). In response to LPS, both domestic pig and warthog TNFα promoters respond significantly to induction ($p<0.05$). However, the domestic pig TNFα promoter responds significantly more strongly than that of the warthog, seen graphically as a steeper induction gradient ($p<0.05$) (Figure 4.5a).

Both promoters respond significantly to TNFα induction ($p<0.05$), but in a similar manner and magnitude (Figure 4.5b). Neither promoter responded to PMA (Figure 4.5c).

5 Functional Analysis of Interspecies Variation in p65

Comparative sequence analysis of the NFκB subunit, p65, identified three differences between the sequence of domestic pig and warthog transcripts (Chapter 3). These three differences occur within the transactivation domains 1 and 2 of p65 and may play a role in determining the activity of NFκB. In particular, the presence of a proline (as opposed to serine) at position 531 in the warthog sequence removes a potential phosphorylation site characterised in human p65 (Bird *et al.* 1997; Wang and Baldwin Jr. 1998). In order to investigate the functional significance of these differences, warthog and domestic pig p65 were expressed to high levels in cell cultures *in vitro*. The activity of NFκB was measured using the firefly luciferase reporter gene downstream from a synthetic promoter containing four tandem copies of the κB consensus binding site.

5.1 Cloning Strategy to Create p65 Expression System

5.1.1 pFLAG-CMV-4 Expression Construct

To enable *in vitro* expression of different p65 genotypes, first-round domestic pig and warthog p65 PCR products from the comparative sequence study (Chapter 3) were diluted to 1:500 in sterile water with 1μl then used as a template in 25μl PCR reaction. This contained 1xPCR buffer, 2mM dNTPs, 2mM $MgCl_2$, 0.7U *High Fidelity* DNA polymerase (Roche Diagnostics) and 20pmol each p65 restriction primer to introduce specific restriction sites (underlined): Forward *Hin*d III (5'-CCA AGC TTG ACC TCT TCC CCC TCA TCT T-3') and reverse *Not* I (5'-GCG CGG CCG CTT AGG AGC TGA TCT GA-3'). Each ~1.6kbp PCR product was run on a 1% agarose gel, excised with a scalpel blade and gel-extracted as previously described (Section 2.2.6).

These restriction sites allowed in-frame insertion of the p65 coding sequence into the multiple cloning site of the pFLAG-CMV-4 vector (Sigma Aldrich). This enables over-expression of p65 driven by the CMV promoter (P_{CMV}). In addition, an N-

terminal eight amino acid FLAG sequence is incorporated into the protein as it is constitutively expressed and translated (Figure 5.1). This FLAG epitope is recognised by an anti-FLAG monoclonal antibody.

Both domestic pig and warthog p65 PCR products and the pFLAG-CMV-4 vector were first digested with *Hin*d III in SuRE/Cut Buffer B and cleaned before being digested with *Not* I in SuRE/Cut Buffer H, this two-stage process was to fulfil different buffer requirements for the two restriction enzymes (Sections 2.3.2 and 2.3.3). The open ends of the vector were dephosphorylated with alkaline phosphatase and the p65 products ligated into it (as described in Sections 2.3.3 and 2.3.4). The resulting pFLAG-p65 plasmids were grown up and sequenced to verify the identity of the inserts and ensure that the cloning was in-frame with the start codon and FLAG sequence.

5.1.2 pNFκB-Luc Reporter Plasmid

In order to quantify the activity of NFκB, the pNFκB-Luc plasmid (BD Biosciences Clontech) was used. This contains four copies of the consensus κB-binding site within a minimal TATA-like promoter (derived from the thymidine kinase promoter of herpes simplex virus) driving expression of firefly luciferase (Figure 5.2). When NFκB binds to the κB elements in the promoter, luciferase is expressed.

5.1.3 Transfections

Vero cells (primate) were transfected (Section 2.4.3) with one of three pFLAG constructs together with the NFκB-Luc reporter construct (Figure 5.3). The three pFLAG constructs contained either domestic pig p65 (Ser531), warthog p65 (Pro531) or no p65 (null) (Figure 5.4). Cells were placed under selection for ten days using 600μg/ml G418. The resulting stably transfected cells were pooled, cultured to increase cell number and stocks frozen down (Section 2.4.2).

Figure 5.1 – pFLAG-CMV-4

The region of the pFLAG-CMV-4 vector (Sigma Aldrich) encoding the eight amino acid FLAG sequence and the multiple cloning site. Warthog or domestic pig p65 was inserted using the *Hin*d III and *Not* I sites.

Diagram taken from Sigma Datasheet for product E1775
pFLAG-CMV-4 Expression Vector

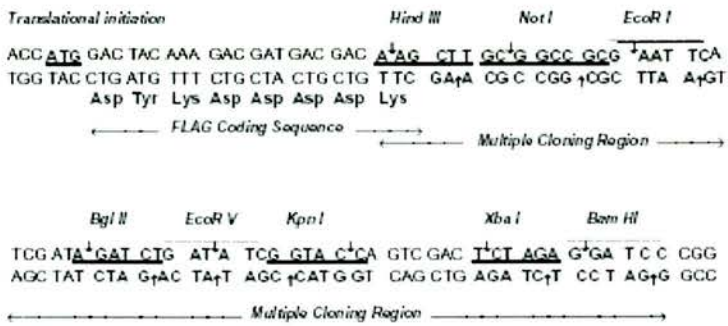

Translational initiation
Hind III
Not I
EcoR I
ACC ATG GAC TAC AAA GAC GAT GAC GAC A↓AG CT T GC↓G GCC GCG ↓AAT TCA
TGG TAC CTG ATG TTT CTG CTA CTG CTG T TC GA↑A CG C CGG ↑CGC TTA A↑GT
Asp Tyr Lys Asp Asp Asp Asp Lys
FLAG Coding Sequence
Multiple Cloning Region
Bgl II
EcoR V
Kpn I
Xba I
Bam HI
TCG ATA↓GAT CTG AT↓A TCG GTA C↓CA GTC GAC T↓CT AGA G↓GA T C C CGG
AGC TAT CTA G↑AC TA↑T AGC ↑CAT G GT CAG CTG AGA TC↑T CC T AG↑G GCC
Multiple Cloning Region

Figure 5.2 – pNFκB-Luc Reporter Construct

The pNFκB-Luc reporter plasmid (BD Biosciences Clontech) was used to monitor the activity of NFκB. It contains the firefly luciferase gene downstream from four tandem copies of the κB consensus sequence (KB_4) fused to a TATA-like promoter (P_{TAL}) region from the Herpes simplex virus thymidine kinase (HSV-TK) promoter. After NFκB binds to the kappa enhancer element (KB_4), transcription is induced and the reporter gene is expressed.

Diagram adapted from 'NF-κB Transcription Reporter Vectors'
CLONTECHniques, July 1998

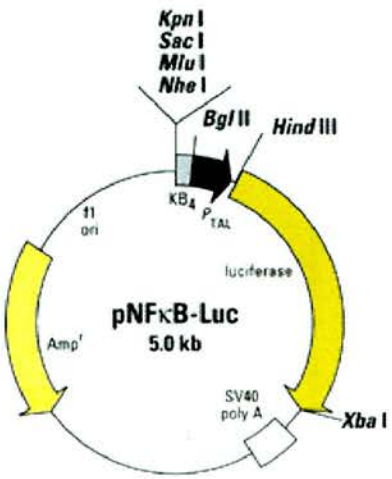

Kpn I
Sac I
Mlu I
Nhe I
Bgl II
Hind III
KB4
PTAL
f1
ori
luciferase
pNFκB-Luc
5.0 kb
Amp'
SV40
poly A
Xba I

Figure 5.3 – p65-expression/Luciferase Reporter Transfections

Diagrammatic representation of co-transfecting a cell with a p65-expression and an NFκB-Luciferase reporter plasmid.

1. The p65-expression and κB-luciferase reporter plasmids are added to the culture medium with a transfection reagent.
2. The plasmids are carried through the cell membrane (transfection).
3. The plasmids linearise and integrate into the host cell DNA.
4. The CMV promoter drives expression of p65.
5. p65 transcripts are translated into protein by the host cell machinery.
6. Exogenous p65 dimerises with endogenous p50 and the newly synthesised NFκB dimer becomes associated with endogenous IκBα.
7. When a proinflammatory signal/inducing agent is added, NFκB is activated via signal-induced phosphorylation and degradation of IκBα.
8. Active NFκB translocates to the nucleus and binds to the four tandem repeats of the κB-binding domain in a minimal promoter driving luciferase expression.
9. Luciferase transcripts are translated into protein and are subsequently quantified using a luminescence assay.

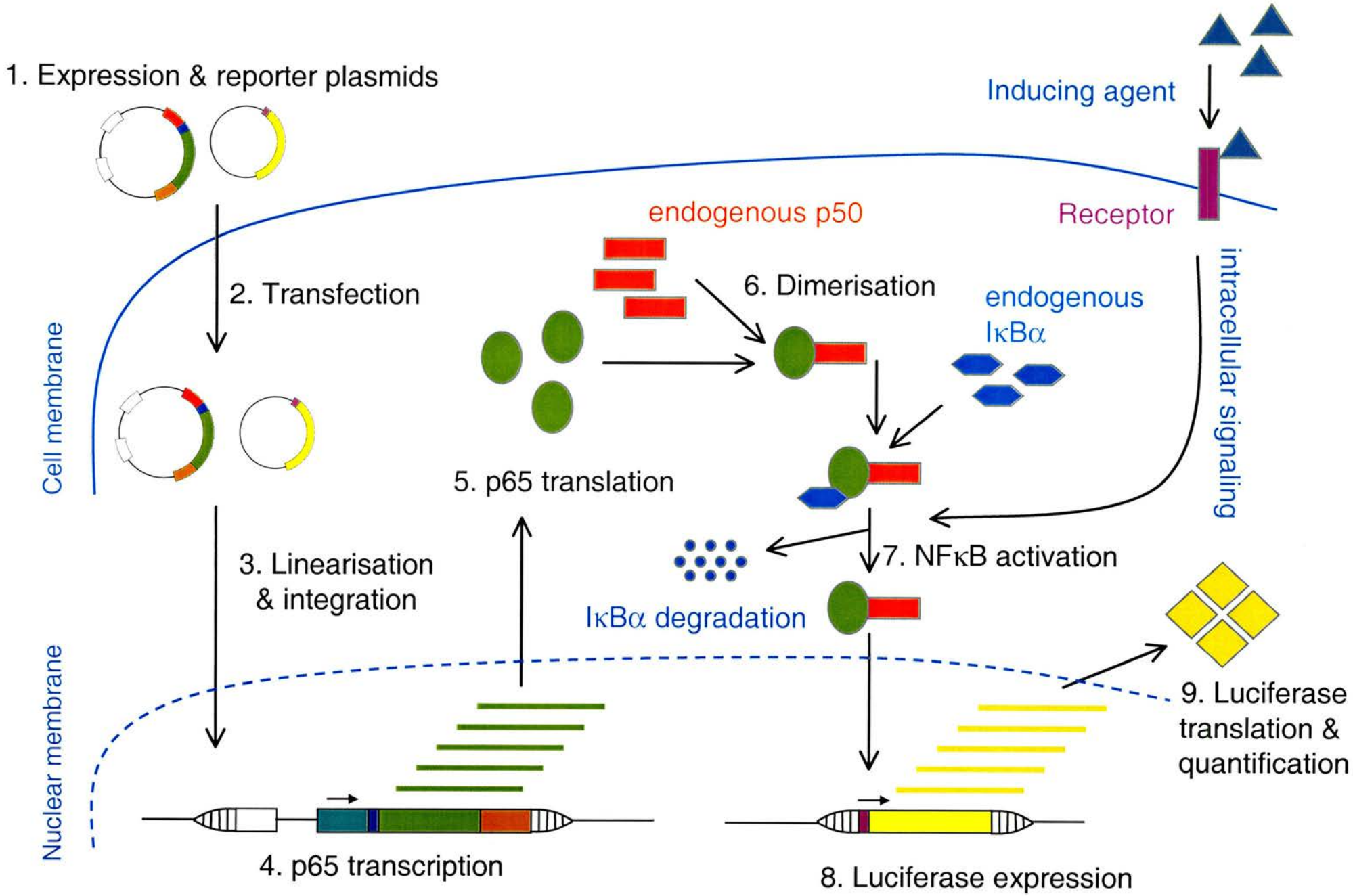

1. Expression & reporter plasmids
Inducing agent
Receptor
endogenous p50
6. Dimerisation
endogenous
IκBα
2. Transfection
Cell membrane
intracellular signaling
5. p65 translation
3. Linearisation
& integration
7. NFκB activation
IκBα degradation
Nuclear membrane
9. Luciferase
translation &
quantification
4. p65 transcription
8. Luciferase expression

Figure 5.4 – pFLAG-p65 Expression and Reporter Constructs

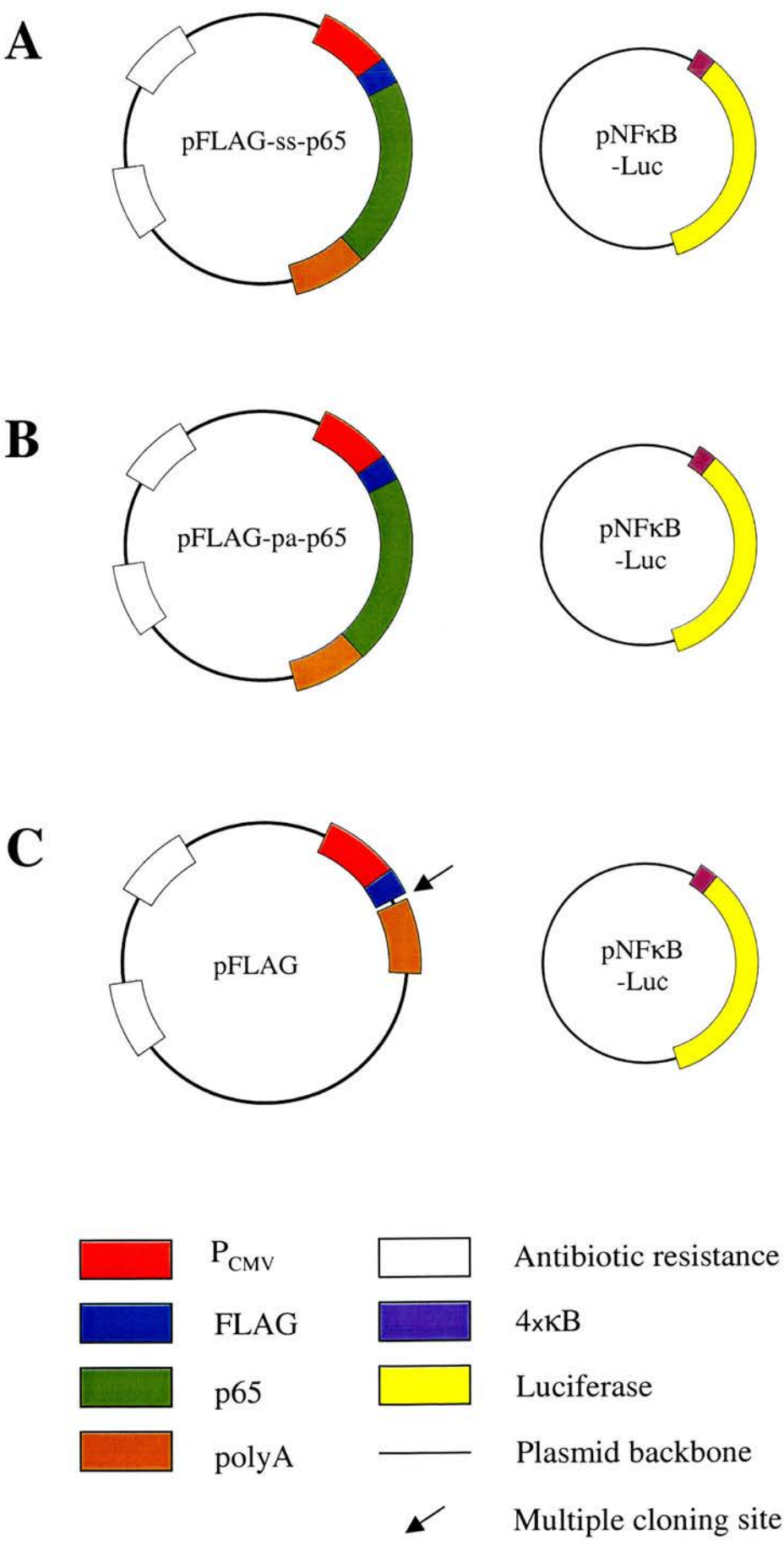
Three co-transfections were carried out to evaluate any difference between the activity of warthog and domestic pig p65.

A Left, the pFLAG-CVM-4 (Sigma Aldrich) construct driving expression of domestic pig (ss) p65; right, the luciferase reporter plasmid, pNFκB-Luc (BD Biosciences Clontech).

B Left, the pFLAG-CVM-4 construct driving expression of warthog (pa) p65; right, the luciferase reporter plasmid, pNFκB-Luc.

C Left, pFLAG-CVM-4 with no insert; right, the luciferase reporter, pNFκB-Luc. This last transfection was used to estimate background NFκB activity and/or any constitutive expression from the 4xκB promoter.

The key below gives details for the colours/symbols used to illustrate the cytomegalovirus promoter (P_{CMV}), the eight amino acid FLAG sequence, the p65 gene, the polyadenylation (polyA) signal derived from human growth hormone (hGH), the G418 and Ampicillin resistance genes, the 4 tandem copies of the κB binding domain in a minimal thymidine kinase promoter (4xκB), the firefly luciferase gene, the multiple cloning site and the plasmid backbone.


A
pFLAG-ss-p65
pNFκB
-Luc
B
pFLAG-pa-p65
pNFκB
-Luc
C
pFLAG
pNFκB
-Luc
P CMV
FLAG
p65
polyA
Antibiotic resistance
4xκB
Luciferase
Plasmid backbone
Multiple cloning site

5.2 Preliminary Induction Studies

Stably transfected cells were plated out into six-well tissue culture plates, so that on the day of the experiment they were ~90% confluent. Eighteen wells were required per stable line (domestic pig, warthog and null p65), to allow for nine experimental conditions carried out in duplicate.

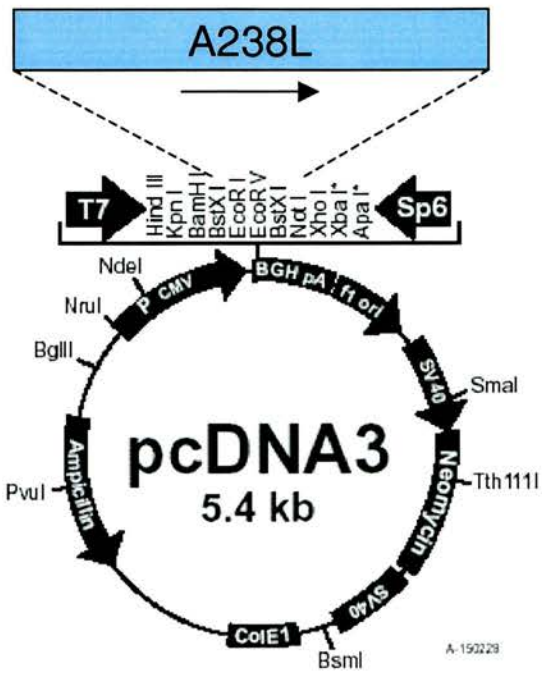
Four wells of each cell line were transfected with a plasmid (pcDNA3, Invitrogen) overexpressing the ASFV-protein A238L from the CMV promoter (Figure 5.5), as previously described by Miskin *et al.* (1998). This construct (subsequently referred to as pA238L) was a generous gift from Dr. Linda Dixon (Institute for Animal Health, Pirbright). The fifth well of each cell line was mock-transfected, using only the lipofectamine reagent (see Section 2.4.3). Eighteen hours post-transfection with pA238L, all cells were treated for six hours (as detailed below) in order to induce NFκB activity and allow accumulation of the luciferase reporter protein. The concentration of each inducing agent used (TNFα and PMA 10ng/ml; LPS 10μg/ml) was based on the comparative TNFα promoter activity results (Chapter 4). It was intended to induce a strong cytokine response, without using an excessive concentration of each agent.

Cells expressing p65	**Treatment**
A. Domestic pig p65 B. Warthog p65 C. Null p65	1. None 2. TNFα (10ng/ml) 3. PMA (10ng/ml) 4. LPS (10μg/ml) 5. Mock transfection (lipofectamine only) 6. pA238L 7. pA238L + TNFα (10ng/ml) 8. pA238L + PMA (10ng/ml) 9. pA238L + LPS (10μg/ml)

Figure 5.5 – pcDNA3 expression A238L (pA238L)

Diagrammatic representation of the pcDNA3 vector (Invitrogen) and the location of the A238L coding sequence inserted into the multiple cloning site. This construct was used to over-express the ASFV protein, A238L, which was driven by the cytomegalovirus promoter (P_{CMV}).

Diagram adapted from pcDNA3 Invitrogen Datasheet (vector no longer available).


A238L
T7
Hind III
Kpn I
BamH I
BstX I
EcoR I
EcoR V
BstX I
Not I
Xho I
Xba I*
Apa I*
Sp6
NdeI
NruI
BglII
P CMV
BGH pA
f1 ori
SV40
SmaI
pcDNA3
5.4 kb
Ampicillin
PvuI
Neomycin
Tth111I
SV40
ColE1
BsmI
A-192229

Following treatment, cells were washed in PBS, harvested in passive lysis buffer (Promega) and assayed for total protein and firefly luciferase (see Sections 2.4.4 and 2.4.5). As described above, all inductions were carried out in duplicate wells and all assays performed in triplicate. Luciferase results were normalised to total protein and once the average results for each well had been determined, the results of the duplicate wells were calculated as an average of the pair (n=2). The raw data from this experiment can be seen in Table 5.1, the results are illustrated in Figure 5.6a. In addition to the experimentally transfected cells, untransfected Vero cells were also used to estimate any background reading from the luminometer. Professional assistance with statistical analyses described below was kindly given by Anthea Springbett (Roslin Institute).

During this experiment, relatively high luciferase activity was seen in uninduced cells, which on some occasions exceeded that detected in induced samples. However, less protein was present in the uninduced cell lysates (~1.1mg/ml) compared to those which had been induced and/or transfected (~1.6mg/ml). Therefore, when luciferase was normalised to total protein, the lower protein value in the uninduced cells caused the ratio to be higher. Possible reasons for this difference in total protein concentration of the lysate will be discussed later.

The presence of domestic pig or warthog p65 caused a significant increase in NFκB activity when the cells were induced with LPS ($p<0.01$), compared to those containing the p65-null construct. However, only warthog p65 caused a significant response to TNFα ($p<0.01$). PMA did not cause any significant induction in cells expressing any p65 construct. The expression of A238L in combination with any agent caused a significant increase in luciferase activity, compared to using the agent alone ($p<0.01$). This was true for cells expressing either domestic pig or warthog p65. The NFκB activity generated by warthog p65 in response to LPS and TNFα was significantly greater than that caused by domestic pig p65, both with and without expression of A238L ($p<0.01$). There was no significant difference between cells expressing warthog or domestic pig p65 when induced with PMA in the presence of A238L (Figure 5.6a). These conclusions are based on Analysis of Variance and

Figure 5.6 – Comparative p65 Activity

A Comparative activity of NFκB in cells stably expressing domestic pig (SS) and warthog (PA) p65. Activity was measured in response to induction with 10ng/ml TNFα, 10ng/ml PMA and 10μg/ml LPS for six hours, both with and without transiently expressed A238L; 'Null' cells contain an empty pFLAG vector. Firefly luciferase readings were normalised to total protein and indicate that NFκB in cells expressing warthog p65 is more transcriptionally active than those expressing domestic pig p65, when induced with TNFα and LPS. Furthermore, A238L appeared to sensitise NFκB to induction.

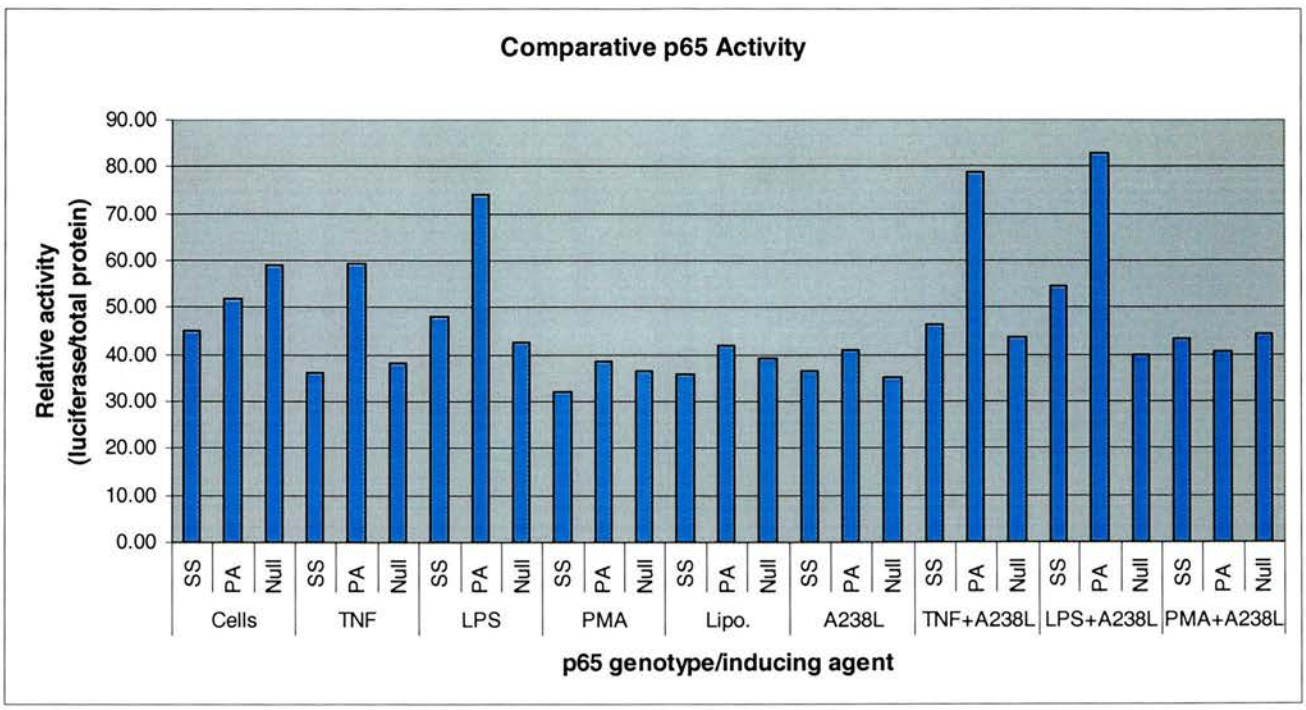
Vero cells stably transfected with the three p65-expressing constructs, were either left uninduced or were induced for 10 minutes using 30ng/ml TNFα. Protein was extracted and run out by SDS PAGE.

Lane	P65	Induced?
M	Marker lane	
1	SS	No
2	PA	No
3	Null	No
4	SS	Yes
5	PA	Yes
6	Null	Yes

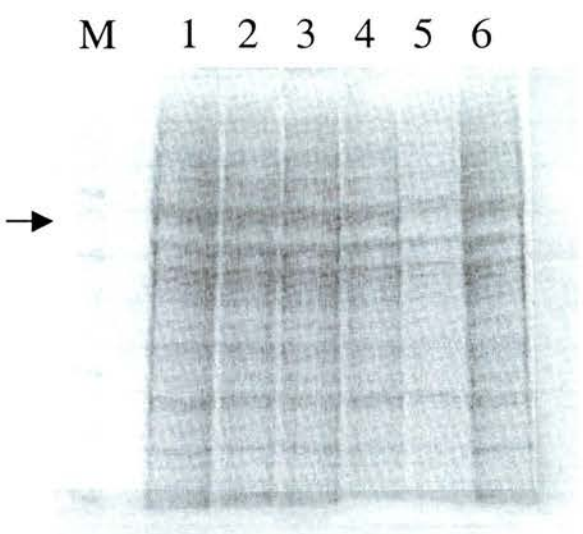
B Total protein visualised using Ponceau S stain; the region of 65kDa has been arrowed. Lane 5 appears to have slightly less protein loaded.

C Western blot probed using a primary anti-p65 polyclonal antibody (LabVision). This blot indicates high levels of p65 in all cells (arrowed), regardless of whether or not they were expressing exogenous p65, indicating the presence of high levels of endogenous p65. The reason behind the lack of a band in lane 5 and the apparent presence of less p65 in the induced cells is unclear. These may be loading artefacts.

A


Comparative p65 Activity
Relative activity (luciferase/total protein)
90.00
80.00
70.00
60.00
50.00
40.00
30.00
20.00
10.00
0.00
SS
PA
Null
Cells
TNF
LPS
PMA
Lipo.
A238L
TNF+A238L
LPS+A238L
PMA+A238L
p65 genotype/inducing agent

B


M 1 2 3 4 5 6

C

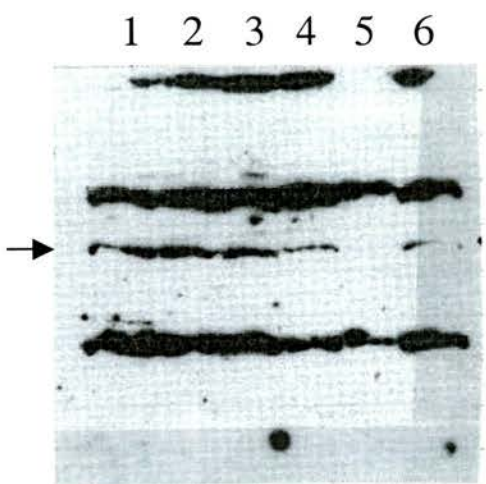

1 2 3 4 5 6

Table 5.1

Spreadsheet of the data obtained from the *in vitro* comparison of domestic pig and warthog p65. NFκB activity was measured by the production of luciferase, driven by a synthetic 4xκB promoter. Each condition was carried out in duplicate and each assay performed in triplicate.

Sample	p65	Treatment	Abs 1	Abs 2	Abs 3	mean Abs	Prot. Conc. (mg/ml)	Luc1	Luc2	Luc3	mean luciferase	mean luc/mear protein
1	dp	cells	2.019	2.105	2.15	2.091	1.197	50	53	49	50.67	42.31
2	dp	cells	1.947	2.027	1.954	1.976	1.069	53	52	48	51.00	47.69
3	wh	cells	2.019	2.138	2.138	2.098	1.205	63	66	66	65.00	53.93
4	wh	cells	2.207	2.134	2.207	2.183	1.299	62	68	64	64.67	49.78
5	null	cells	1.937	1.958	2.018	1.971	1.064	53	58	56	55.67	52.33
6	null	cells	1.793	1.771	1.969	1.844	0.923	61	61	60	60.67	65.73
7	dp	tnf	2.548	2.77	2.703	2.674	1.845	66	67	68	67.00	36.32
8	dp	tnf	2.64	2.593	2.593	2.609	1.772	69	60	62	63.67	35.92
9	wh	tnf	2.418	2.514	2.417	2.450	1.596	91	93	92	92.00	57.66
10	wh	tnf	2.389	2.447	2.447	2.428	1.571	98	98	90	95.33	60.68
11	null	tnf	2.417	2.514	2.417	2.449	1.595	57	57	61	58.33	36.57
12	null	tnf	2.442	2.483	2.483	2.469	1.617	66	66	62	64.67	39.98
13	dp	lps	2.37	2.337	2.441	2.383	1.521	84	81	83	82.67	54.34
14	dp	lps		2.58	2.706	2.643	1.810	77	73	77	75.67	41.79
15	wh	lps	2.483	2.638	2.403	2.508	1.660	136	146	131	137.67	82.91
16	wh	lps	2.575	2.664	2.616	2.618	1.783	124	112	111	115.67	64.87
17	null	lps	2.386	2.575	2.616	2.526	1.680	77	67	77	73.67	43.85
18	null	lps	2.538	2.839	2.662	2.680	1.851	78	74	76	76.00	41.05
19	dp	pma	2.362	2.772	2.772	2.635	1.802	65	56	57	59.33	32.93
20	dp	pma	2.535	2.48	2.757	2.591	1.752	58	54	51	54.33	31.01
21	wh	pma	2.455	2.564	2.631	2.550	1.707	67	68	60	65.00	38.08
22	wh	pma	2.348	2.533	2.631	2.504	1.656	63	62	67	64.00	38.65
23	null	pma	2.247	2.455	2.506	2.403	1.543	59	56	63	59.33	38.44
24	null	pma	2.388	2.513	2.417	2.439	1.584	52	58	54	54.67	34.51
25	dp	lipo	2.417	2.551	2.388	2.452	1.598	59	56	59	58.00	36.29
26	dp	lipo	2.388	2.446	2.388	2.407	1.549	58	50	55	54.33	35.09
27	wh	lipo	2.362	2.592	2.416	2.457	1.603	74	71	70	71.67	44.70
28	wh	lipo	2.487	2.349	2.349	2.395	1.535	61	60	58	59.67	38.87
29	null	lipo	2.277	2.369	2.435	2.360	1.496	60	60	62	60.67	40.54
30	null	lipo	2.329	2.435	2.435	2.400	1.540	56	56	63	58.33	37.88
31	dp	a238l	2.435	2.577	2.69	2.567	1.726	55	62	56	57.67	33.40
32	dp	a238l	2.287	2.337	2.375	2.333	1.466	56	66	53	58.33	39.79
33	wh	a238l	2.513	2.356	2.462	2.444	1.589	65	61	58	61.33	38.60
34	wh	a238l	2.084	2.762	2.375	2.407	1.548	63	66	71	66.67	43.06
35	null	a238l	2.375	2.717	2.812	2.635	1.801	66	61	55	60.67	33.68
36	null	a238l	2.402	2.435	2.402	2.413	1.555	55	59	57	57.00	36.66
37	dp	tnf-a238l	2.646	2.402	2.507	2.518	1.672	73	74	74	73.67	44.06
38	dp	tnf-a238l	2.247	2.206	2.646	2.366	1.503	76	73	70	73.00	48.57
39	wh	tnf-a238l	2.226	2.469	2.434	2.376	1.514	130	127	119	125.33	82.77
40	wh	tnf-a238l	2.314	2.337	2.195	2.282	1.409	104	106	105	105.00	74.50
41	null	tnf-a238l	2.479	2.446	2.638	2.521	1.675	69	73	74	72.00	42.99
42	null	tnf-a238l	2.25	2.446	2.416	2.371	1.508	65	69	65	66.33	43.99
43	dp	lps-a238l	2.337	2.292	2.416	2.348	1.483	79	86	80	81.67	55.07
44	dp	lps-a238l	2.279	2.336	2.306	2.307	1.437	74	82	77	77.67	54.04
45	wh	lps-a238l	2.336	2.58	2.441	2.452	1.599	143	127	137	135.67	84.87
46	wh	lps-a238l	2.306	2.336	2.441	2.361	1.497	126	122	116	121.33	81.04
47	null	lps-a238l	2.336	2.483	2.578	2.466	1.613	62	66	65	64.33	39.87
48	null	lps-a238l	2.412	2.538	2.44	2.463	1.611	65	66	61	64.00	39.73
49	dp	pma-a238l	2.236	2.44	2.44	2.372	1.509	65	63	61	63.00	41.74
50	dp	pma-a238l	2.254	2.273	2.44	2.322	1.454	65	68	64	65.67	45.16
51	wh	pma-a238l	2.712	2.662	2.77	2.715	1.890	78	82	73	77.67	41.09
52	wh	pma-a238l	2.431	2.455	2.564	2.483	1.633	66	67	63	65.33	40.01
53	null	pma-a238l	2.107	2.408	2.367	2.294	1.423	70	65	58	64.33	45.22
54	null	pma-a238l	2.348	2.329	2.311	2.329	1.462	66	59	64	63.00	43.09

comparison of means for cells expressing domestic pig p65, warthog p65 and those containing the null-p65 construct. From the ANOVA it was found that there is no interaction between the effect of A238L and the effects of p65 genotype and inducing agent. This means that during this study, the effect of adding A238L is consistent across the other treatment combinations.

The lack of response from cells containing either p65 construct to PMA may reflect a similar situation to that observed in the TNFα promoter study, where no induction was seen in response to increasing concentrations of PMA (section 4.4.1.4). The presence of transiently expressed A238L alone had little effect on NFκB activity. However, it significantly increased the response of NFκB to inducing agents.

Aliquots of the three sets of transfected cells described above were also cultured in T-75 flasks to provide sufficient material for protein extraction. These cells were either left unstimulated or stimulated with 30ng/ml TNFα for ten minutes, to provide an option for measuring for any signal-induced change in phosphorylation status of p65 that may occur between the constructs as described by Wang and Baldwin (1998). These cells were then washed with PBS, trypsinised from the flasks and lysed in homogenisation buffer containing antiproteases (for constituents see Table 2.4). Total protein was extracted and run as a Western blot, probing with an anti-p65 polyclonal antibody (LabVision). The Western blot illustrated high levels of p65 in all cells, regardless of whether they were expressing exogenous p65, indicating the presence of high levels of endogenous p65 (Figures 5.6b and 5.6c). It was considered that such high levels of endogenous p65 would be likely to obscure any subsequent attempts to measure the phosphorylation status of exogenous p65.

5.3 Summary - Comparative p65 Activity

p65-expression constructs have been developed to investigate functionality attached to sequence differences identified between the domestic pig and warthog transcripts. This pilot study indicates that when induced with TNFα and LPS, NFκB appears to be significantly more transcriptionally active in the presence of warthog p65 than

domestic pig p65 ($p<0.01$). In addition, the presence of A238L significantly increases the activity of NFκB in induced cells expressing either warthog or domestic pig p65 ($p<0.01$). Western blot analysis indicated high background levels of endogenous p65 in the transfected cells, which may obscure clear detection of differences between these experimental constructs at the protein level (by physical and/or visual interference). Nevertheless, these results raise several intriguing questions regarding the function of these polymorphisms in p65, which conflict with the initial hypothesis that warthog p65 should be less transcriptionally active. Furthermore, the presence of transiently expressed A238L appeared to sensitise NFκB to induction in this system, despite its previously demonstrated role as an NFκB 'super-repressor' *in vitro* (Powell *et al.* 1996; Tait *et al.* 2000).

6 Discussion

6.1 Introduction and Context

African swine fever virus (ASFV) poses one of the greatest threats to pig farming worldwide (Pan and Hess 1984; Monath 1986; Fenner *et al.* 1993). It is a highly infectious pathogen which causes devastating haemorrhagic disease of domestic pigs and Eurasian wild boar (*Sus scrofa*) (Oura *et al.* 1998b; Takamatsu *et al.* 1999). In contrast to the severe pathology seen in domestic pigs, the native African suids (warthogs and bushpigs) are asymptomatic carriers of infection which suffer only a mild subclinical form of the disease from which they rapidly recover (Montgomery 1921; DeTray 1963; Mansvelt 1963; Thomson *et al.* 1980; Anderson *et al.* 1998; Oura *et al.* 1998a).

To investigate the genetic basis of host immune response to a pathogen, there are two principle approaches which may be adopted: (1) A genome-wide approach, for example using microarray technology, to identify genes whose expression is affected by infection and (2) a candidate gene approach, which involves selecting previously identified genes for further investigation. A cDNA microarray has recently been developed to assess changes in gene expression of domestic pig macrophages in response to ASFV infection by Dr. Stewart Lowden (University of Edinburgh), Dr. Linda Dixon (IAH Pirbright) and Prof. Alan Archibald (Roslin Institute). However, the results of this analysis have not yet been published. As this technology was not available at the time of initiation of this study, a candidate gene approach was adopted. The TNFα promoter and coding sequence of six other genes were selected on the basis of existing evidence indicating that they were cellular targets for ASFV proteins and/or involved in mediating ASFV pathophysiology.

An initial hypothesis examined the potential role of tumour necrosis factor alpha (TNFα) in the ASFV disease process. During ASFV infection of domestic pigs, a massive increase is seen in levels of the proinflammatory cytokine, TNFα (Gómez del Moral *et al.* 1999; Carrasco *et al.* 2002; Salguero *et al.* 2002; Gil *et al.* 2003).

High levels of TNFα expression have previously been associated with a range of other viruses which cause haemorrhagic fever and lymphopaenia (Feldmann *et al.* 1996; Marianneau *et al.* 1998; Marta *et al.* 1999; Villinger *et al.* 1999; Geisbert *et al.* 2000; Sato *et al.* 2000; Azeredo *et al.* 2001; Kurane and Takasaki 2001; Baize *et al.* 2002). This raises the possibility that TNFα may be involved in causing the severe pathology seen during ASFV infection of domestic pigs. As TNFα production is primarily driven by its *cis*-acting promoter (Kuhnert *et al.* 1991; Haudek *et al.* 1998), it was hypothesised that variation in the TNFα promoter sequence may have a significant impact on expression.

A second hypothesis investigated here involved the cellular targets of ASFV proteins A238L and p54. *In vitro* evidence suggests that the ASFV-encoded protein, A238L, interacts with key elements of porcine NFAT and NFκB signalling pathways to effectively modulate the host immune response (Powell *et al.* 1996; Miskin *et al.* 1998; Miskin *et al.* 2000; Tait *et al.* 2000). In addition, the ASFV protein, p54, is involved in attachment of virus particles to the microtubule motor complex, cytoplasmic dynein. This results in the perinuclear localisation of virus particles and may represent an important stage in the infection process (Alonso *et al.* 2001). It was therefore hypothesised that sequence differences in proteins targeted and/or mimicked by ASFV proteins A238L and p54, may be involved in determining the different pathophysiological outcomes of ASFV infection in suid species of varying susceptibility.

6.2 TNFα Promoter

6.2.1 Experimental Approach

TNFα expression is driven primarily by a *cis*-acting promoter which lies in the ~1.6kbp intergenic sequence separating the start codon of TNFα from the upstream gene encoding TNFβ (Kuhnert *et al.* 1991; Solinas *et al.* 1992; Haudek *et al.* 1998). This region was amplified from genomic DNA by PCR from the ASFV-susceptible domestic pig, resistant warthog and phenotypically unknown babirusa. The

amplicons were ligated into a sequencing vector and sequenced in all three species. Comparative transcription factor binding site analysis was carried out using components of the Genomatix software suite. This software suggested variation in a number of putative transcription factor binding sites and modules. To investigate functionality associated with sequence differences, the warthog and domestic pig TNFα promoters where cloned into firefly luciferase reporter plasmids, which were subsequently used to create two novel dual-reporter plasmids. These encoded a control renilla luciferase cassette in addition to the experimental, TNFα promoter-driven firefly luciferase cassette. Dual-reporters were transfected into murine Hepa 1 cells in tissue culture and stable colonies selected by the use of zeocin. In order to assess activity of the TNFα promoters, these cell lines were induced using three agents (LPS, TNFα and PMA) to elicit an immune-type response. Differences in activity between the warthog and domestic pig TNFα promoter were measured by comparing firefly luciferase normalised to control renilla luciferase activity.

6.2.2 Comparative Sequencing

The TNF 'intergenic region' is 1600bp in length in the domestic pig, 1592bp in the warthog and 1599bp in the babirusa. Thirty sequence differences were identified between the warthog and domestic pig. In addition, the warthog sequence has two single base deletions and one six base pair deletion. MatInspector^{TM} detected 170 potential binding sites in the domestic pig (Appendix 1) and 173 in the warthog (Appendix 2). Of these, 149 are conserved between the species; however there are 21 sites present in only the domestic pig and 24 sites only in the warthog (Tables 3.8 and 3.9). Most of these results were additional copies of sites already present in multiple locations within the promoters of both species. It is difficult to predict the significance of so many putative sites and to suggest the importance of copy number and location of a particular site. Therefore, ModelInspector^{TM} was used as a second, more robust tool to detect modules of two or more transcription factor binding sites. Ten putative modules were identified in the domestic pig and 17 in the warthog sequence (Table 3.10). Only one module is unique to the domestic pig sequence, whereas the warthog contains eight putative modules not present in the domestic pig

sequence. Although the proximal promoters are highly conserved between the two species, within a short region of distal sequence the warthog sequence encodes eight predicted overlapping modules. The same region in the domestic pig contains only two putative modules (Figure 3.12).

During truncation experiments involving the human TNFα promoter, Haudek *et al.* (1998) found evidence for a repressor/negative-regulatory element in the distal region of the promoter. When they examined only the proximal promoter (600bp upstream of the transcription start point), its activity was approximately three-fold greater in response to LPS induction than when using the entire promoter. It is therefore possible that these distal differences between warthog and domestic pig sequences may be involved in negative-regulation. Although the distance of these distal differences from the transcription start point (<1kbp) may preclude them from having any direct transcriptional function, they could still act in a number of ways. At the simplest level, a gathering of transcriptional machinery on the distal promoter may inhibit the binding of 'genuine' transcription factors by some form of long-range interference. Alternatively, transcription factors binding to this region may in turn recruit cofactors which alter the histone structure and transcriptional potential of the DNA. Histone modifications have been previously shown to modulate the activity of the TNFα promoter (Lee *et al.* 2003) and if these were negative-regulatory (e.g. histone deacetylation) they could explain the findings of Haudek *et al.* (1998) in the human TNFα promoter.

As the DNA sequence itself has a profound effect on chromatin structure, the existence of sequence differences may play a significant role in mediating transcriptional activity. By altering the precise position of a nucleosome and how the DNA is wrapped around it (both translational and rotational positioning) will determine the accessibility of transcription factor binding sites on either strand of the duplex. Even if a particular sequence difference is not directly involved in a binding site it may still mediate transcriptional activity by this indirect mechanism (Beato and Eisfeld 1997). To investigate whether sequence differences between the domestic pig and warthog TNFα promoters affect nucleosome organisation, the

position of nucleosomes can be mapped using the monomer extension assay. This technique tests candidate sequence for regions which are protected from DNAse (micrococcal nuclease) hypersensitivity by being bound to a nucleosome (Yenidunya *et al.* 1994; Whitelaw 2000). Testing for DNAse I hypersensitivity is also a useful technique. DNAse I hypersensitive sites (DH-sites) are usually associated with regions of DNA which are associated with complexes, often transcription factors, and are therefore considered indicative of functional importance. Initial DNAse I hypersensitivity studies of the porcine TNF gene locus identified two DH-sites in the macrophage TNFα gene (Kuhnert *et al.* 1992). One DH-site was located close to the TATA-box in the proximal promoter and the other towards the 3' end of intron three. Neither of these sites was observed in non-TNFα-expressing cells (testicular and kidney fibroblasts). Not only does this study reinforce the functional significance of chromatin structure on gene expression, but it also illustrates that functionally important responsive elements may be located outwith the *cis*-acting promoter. Kwon *et al.* (1996) also identified an important regulatory element and DH-site in the 3' flanking region of the TNFα gene in astrocytes (located within 500bp downstream of the polyadenylation signal). Other examples of this phenomenon are the location of an important erythroid-specific regulatory element of the human alpha-globin gene in intron five of an upstream gene *-14* (Vyas *et al.* 1995) and the existence of a TNF responsive element within intron two of the manganous superoxide dismutase (Mn-SOD) gene (Guo *et al.* 2003). It is important to acknowledge that the effects mediated by such 'non-promoter' sites will not have been detected during the comparative analysis of the TNF intergenic region undertaken within this thesis.

The close proximity of TNFα and TNFβ cannot be overlooked, as a distance of only 1.6kbp between two genes is unusual. TNFα and TNFβ (which are predominantly expressed by macrophages and lymphocytes respectively) are structurally and functionally related cytokines, performing overlapping roles and acting on the same cell-surface receptors (Kuhnert *et al.* 1992). Although there have been no reports of a direct transcriptional relationship between these genes, they must at the very least accommodate each other within the short region of the genome in which they are

both located. However, their functional similarity and close proximity may suggest a far more intimate relationship. Furthermore, the intergenic region between them is highly conserved, which may be indicative of functionality. Curiously, the Javan warty pig (*Sus verrucosus*), which was the only suid to show any significant variation from this by having a 308bp insertion in the intergenic region, died as a neonate (Appendix 3). However, as this rare sample was the only available representative of this species, the insertion could not be investigated in additional Javan warty pigs. It could also be speculated that this variable distal region of the TNFα promoter, which lies within close proximity to the polyadenylation signal of TNFβ, could influence transcription of TNFβ, TNFα or both TNF genes. For example, transcription factor binding sites in the 'distal promoter of TNFα' may in fact interrupt TNFβ transcription, acting as an insulator to prevent read-through or movement of the transcription machinery towards TNFα. A relaxation of inhibition in this region could potentially encourage tandem expression as if the region is active and transcription factors and polymerase components have already been recruited to the immediate vicinity, the probability of successful transcription in adjacent genes would be increased. Alternatively, these binding sites may be genuine response elements for TNFβ, but they happen to be located in its 3' flanking region. Therefore, in order to gain a thorough understanding of TNFα expression, it may be necessary to consider the entire ~10kbp TNF locus as a single extended functional unit.

It is important to recognise that while they offer a valuable initial analysis, MatInspector and ModelInspector are designed primarily to identify mouse and human transcription factor binding sites and modules. Although many of these are highly conserved across mammals, the suid sites may not be a perfect match for those of humans and/or mice and may be either missed by the software or scored incorrectly and rejected. Furthermore, there is huge potential for the generation of false-positive sites, especially those with short or repetitive recognition sequences. However, to further this *in silico* analysis, the Genomatix software suite contains a range of programs which can be subscribed to. For example, human, mouse, and rat TNFα promoter sequences could be retrieved from the public database using

Eldorado™. These sequences could then be entered into FrameWorker™ to create a set of consensus models of how transcription factor binding sites and modules occur within the mammalian TNFα promoter. ModelInspector could subsequently be used to test query TNFα promoter sequences for these user-defined FrameWorker models. Strong patterns of evolutionary conservation would become evident, as well as highlighting sequences, or regions of sequences, which appear significantly removed from the evolutionary norm. Furthermore, these models could also be used to search against a database of other promoters. This may identify potentially functionally related genes, whose promoters include regions similar to those defined by the TNFα promoter models. By improving the reliability and resolution of the *in silico* analysis, the selection of sites worthy of further *in vitro* or *in vivo* analysis would present a less daunting task.

Once individual base differences have been identified as potentially significant, further statistical analysis could be used to assess if these positions appear to be under positive selection. Techniques have been successfully employed to confirm positive selection for an additional NFAT-binding site in the human interleukin 4 (IL4) promoter, compared to that of the other great apes (Rockman *et al.* 2003). This method measured the relative pressure on a particular C→T polymorphism in different human populations by comparing its distribution to that of 18 unlinked single nucleotide polymorphisms (SNPs). However, to undertake this type of analysis in suids would require selection and sequencing of a similar panel of 'unlinked' SNPs.

6.2.3 Comparative Activity

Dual-luciferase reporter plasmids were developed to investigate if sequence differences between the domestic pig and warthog TNFα promoter have a significant impact on function (Chapter 4). During these *in vitro* experiments, the domestic pig promoter appeared to be inherently more active than the warthog and also responded significantly more strongly to LPS (Figure 4.5). Both promoters responded in a similar manner and magnitude to increasing levels of TNFα, although they started at

significantly different basal levels. Neither promoter responded significantly to PMA. These *in vitro* findings may reflect an important functional distinction between warthog and domestic pig TNFα expression *in vivo*. The lower levels of activity seen in the warthog TNFα promoter may be further evidence for the existence of a negative-regulatory domain in the distal sequence, which is absent (or less active) in the domestic pig.

6.2.3.1 Experimental Considerations

From the outset, it is important to acknowledge that in all *in vitro* experiments involving tissue culture, transfection, induction and reporter gene analysis there are many factors which remove it from the *in vivo* situation. These issues are inherent to the nature of such experiments and although every effort can be taken to minimise these effects, it must be understood that the objective was to create a controlled, albeit experimental, system to investigate a hypothesis, not to perfectly mirror the *in vivo* environment. A discussion of these issues is presented below.

During this study, the experimental TNFα promoters were not driving transcription of the TNFα mRNA transcript, but that of luciferase. It is widely recognised that many mRNA transcripts, both coding and non-coding regions, contain inherent sites or domains which influence their own transcription and subsequent processing and regulation. For example, such domains within the transcript may be involved in determining the stability of the mRNA transcript and influence how much protein is translated from it in a transcriptionally-independent manner (Rousseau *et al.* 2003). Furthermore, the synthesised protein frequently performs a range of autoregulatory functions, which may implement negative or positive feedback loops. Although this lack of post-transcriptional and post-translational levels of control may remove this system from *in vivo* TNFα production, this does not invalidate the aim of the experiment, which was to simply compare the activity of two promoter sequences. It should also be noted that the assay used during this experiment relied on measuring the accumulation of the luciferase reporter protein over a period of hours. While this situation is clearly removed from the rapid cytokine response pathway which occurs

within minutes *in vivo*, this approach was designed to allow the relative activity of the two promoters to be compared in a controlled manner.

It is important to identify factors which affect the structure of the promoter sequences when evaluating their activity. In the transfected cells, the TNFα promoters are out of their usual chromosomal context and the assay will therefore be primarily assessing *cis*-acting effects rather than any *trans* effects attributed to response elements or genes located in the surrounding sequence. In addition, the chromatin structure of the transgene sequence may be very different to normal. This can have a profound effect on transcriptional activity, as has previously been demonstrated in the TNFα promoter (Lee *et al.* 2003). To normalise for differences caused by insertion site (both chromosomal context and chromatin rearrangement), stably transfected cells were pooled. However, populations of pooled colonies can change over time, for example if one population displays growth advantage due to the location of its transgene. For this reason, individual colonies of stably transfected cell lines can be isolated and grown up to produce a range of clonal lines with stable characteristics. The average effect of a given treatment on a range of these clonal lines would provide a more accurate system for long-term investigation.

Cell type and species are also important considerations when designing an *in vitro* experiment. Although a number of cell types can produce TNFα, it is primarily the role of macrophages to do so. Although hepatocytes/liver-derived Hepa 1 cells respond to TNFα, they do not naturally express TNFα. In order to address this issue, a macrophage-derived cell line could be used instead. For example, Haudek *et al.* (1998) used mouse macrophage line, RAW 264.7, to investigate differences in activity between the human and baboon TNFα promoters. Indeed, a porcine macrophage line has recently been established (IPAM), although there has been some discussion as to whether or not this is infected with porcine circovirus 2 (PCV2) (Linda Dixon, personal communication). The possible use of porcine cell lines raises additional questions regarding species compatibility of the cell line. Clearly, the cellular machinery in a primate or murine cell is not an ideal match for suid transgenes. However, while the use of porcine (domestic pig) cells lines has

several distinct advantages including compatibility and the potential for porcine virus challenge, during comparative experiments involving warthog genes this may bias the results towards the domestic pig transcript/protein. Therefore, both an appropriate cell type and species should ideally be selected. To fulfil these criteria, a possible candidate line might be one derived from babirusa macrophages, which would represent an 'independent' suid line of the correct cell type. However, no such line currently exists and any assumption of domestic pig-warthog species bias would be purely speculative. Furthermore, regardless of species compatibility, the use of a macrophage cell line has its own drawbacks. Macrophages are terminally differentiated cells and any line derived from these cells may display unnatural characteristics. If primary cells (for example alveolar or peripheral blood macrophages) were isolated and used as an alternative, these could only be transiently transfected and cultured for a short period of time. Although this might suffice for a single experiment, many cells would be required to provide an adequate number for different experiments and replicates. Each batch of transfections would display different characteristics and expression profiles, making normalisation and comparison of results difficult. Moreover, during transient transfections, episomal DNA has only limited chromatin structure and may display an atypical transcriptional profile.

The choice of control promoter is another key consideration. It is widely accepted that no control promoter is truly constant across treatments or functions completely independently from all effects on the cell. Indeed, most constitutive promoters are viral in origin and as such have evolved to respond to host-cell transcription factors and machinery. In this study, the thymidine kinase promoter (P_{TK}) was selected as it is small, relatively uninfluenced by outside factors on the cell and is less susceptible to a gradual decay in activity seen in other promoters, e.g. P_{CMV} (as advised by Promega Technical Services). Although the results presented in this thesis indicate that renilla luciferase activity was remarkably stable (Figure 4.4), some researchers have abandoned renilla controls altogether (Perkins, ND, personal communication). They have chosen instead to normalise by total protein to account for cell number. However, this approach presents complications as different experiments have different numbers of cells which display different copy number of plasmid per cell.

Indeed, even if a Southern blot was performed to demonstrate copy number, it would be impossible to determine how many of the copies present were in fact active. In this study, the use of a dual-reporter approach was found to provide a robust self-normalising system. As results are always a ratio of firefly to renilla luciferase at the per-plasmid level, all other factors should remain constant. Furthermore, should interference, read-through, P_{TK}-activation or any other unaccounted-for-effects have been taking place, we can assume that these were the same across all experiments and can therefore be ignored.

The choice of inducing agents and range of concentrations was based on a number of previous experiments of a similar nature (de Martin *et al.* 1993; Powell *et al.* 1996; Abu-Amer *et al.* 1997; Haudek *et al.* 1998; Revilla *et al.* 1998; Wang and Baldwin Jr. 1998; Tait *et al.* 2000; Wang *et al.* 2000). Although the concentration range of LPS and TNFα gave a strong response which could be differentiated between species, a more detailed experiment could be carried out to measure the response phase at lower concentrations. This may enable a more accurate comparison of activity between the two promoters, rather than relying primarily on the plateau phase.

PMA (analogous to endogenous diacylglycerol, DAG) has previously been used to induce a protein kinase C (PKC) response in many cell types including a murine macrophage line (Haudek *et al.* 1998), human T cell lines (Falvo *et al.* 2000) and primary porcine endothelial cell cultures (Vallée *et al.* 2001). However, it failed to produce any response from the exogenous TNFα promoters during these experiments. The lack of response seen in Hepa 1 cells may have occurred for a number of reasons. There could be incompatibility issues causing this cell type or cell line not to respond to PMA, for example due to a lack of required receptors or second messengers or that a higher concentration of PMA is required to elicit a response. Alternatively, the TNFα promoter sequences under investigation may not include the necessary response elements, possibly due to species incompatibility or that they exist outwith the *cis*-acting promoter. A further option which could be investigated would be to combine PMA with an ionophore (e.g. ionomycin) to

address the possibility that a lack of calcium mobilisation may be reducing the level of induction. When using PMA to mimic DAG, the production of IP_3 is by-passed (see Section 4.4.3). However, ionomycin mobilises Ca^{2+} release from intracellular stores in a similar manner to IP_3 and causes activation of PKC and calcium-dependent transcription factors such as NFAT (Morgan and Jacob 1994; Schaefer 2004). Used in combination, PMA and ionomycin may better replicate the DAG/IP_3-mediated response.

6.2.3.2 Interpretation of Results

If these *in vitro* findings reflect genuine differences in the activity of the TNFα promoter, this variation may have evolved for a number of reasons. It is tempting to speculate that lower levels of proinflammatory cytokine expression in warthogs may reduce the chance of haemorrhagic fever, endothelial cell dysfunction, hypergammaglobinaemia, disseminated intravascular coagulation (DIC) and lymphocyte apoptosis. However, it is also possible that a strong TNFα response is essential for the survival of domestic pigs following exposure to pathogens to which the warthog may exhibit greater levels of susceptibility. The warthog may have evolved to exhibit this less intense response as a result of ASFV (and/or other selection pressures) or may demonstrate alternate resistance mechanisms which necessitate only a moderate TNFα response. Alternatively, the warthog sequence may have simply 'drifted' due to a lack of positive selection pressure. It is conceivable that a low level TNFα expression phenotype may benefit the warthog in combating continued chronic insult such as low grade viral/bacterial infection, parasite or tick infestation. The domestic pig on the other hand may be better-adapted to producing a strong inflammatory response to rapidly clear infection. Neither of these strategies is necessarily perfect for all eventualities. However, it may demonstrate a clear evolutionary distinction between these species.

A key question which arises on consideration of this dichotomy is whether a domestic pig expressing TNFα in the same manner as a warthog would be less likely to die from haemorrhagic fever. Equally, given such circumstances, would the

domestic pig rapidly succumb to an alternative disease to which this response has evolved as a protective mechanism?

As mentioned previously, it is important to acknowledge that this study has not taken into account higher levels of regulation (i.e. post-transcriptional, translational and factors present in the native cell). It may therefore be true that regardless of any differences in translational activity observed *in vitro*, the warthog and domestic pig may in fact produce the same amount of protein *in vivo* due to higher, as yet unknown, compensatory/complimentary mechanisms.

6.2.4 Future TNFα Promoter Work

There is much potential for future expansion of this work now that a novel dual-luciferase reporter system has been successfully developed. Indeed, this system could be utilised to test functional differences between any two (or more) candidate promoters. In addition to repeating the experiment described here, further work in comparative suid TNFα promoter studies may include the following three approaches:

The *first* approach involves an in-depth investigation into the precise mechanisms underlying the difference in TNFα promoter activity. Thus the current study could be expanded to investigate:

- Additional inducing agents
- More focussed dose response curves
- Different cell lines (both species and cell type)
- Clonal populations of transfected cells (rather that pooled colonies)
- Co-transfection with constructs expressing viral proteins (e.g. A238L)
- Anti-TNFα and other drugs
- Viral infection *in vitro*
- Chromatin and histone arrangement of the two promoters

Mutation and/or truncation analysis of the TNFα promoter may also be carried out to investigate the role of specific regions and individual bases within the sequence. This approach could be broadened to include intronic sequence and even the entire TNF locus. This may highlight elements governing transcription of either TNFα or TNFβ or both genes located outwith the intergenic region.

It may also be possible to develop an alternative dual-reporter system comprising both the domestic pig and warthog TNFα promoters driving the two different luciferases in a single vector. This would enable a direct comparison of their activity, although it is conceivable that there may be some risk of interference and the sequestration of transcription factors and other cellular resources. If the luciferase genes were swapped in parallel experiments (similar to a microarray experiment dye-swap) the results should eliminate, or at least identify, idiosyncratic interactions of test promoters and luciferase genes. However, to achieve this goal would require a complex cloning strategy due to the high number of restriction sites common to both the promoter sequences and the various plasmids required during the procedure.

A *second* approach is to proceed with the information established in this study and to apply this knowledge in downstream experiments. The results indicate that the warthog TNFα promoter is significantly less active than that of the domestic pig, an observation which may be of functional importance *in vivo*. Therefore, therapeutic questions could be asked both *in vitro* and *in vivo*. For example, do anti-TNF/anti-inflammatory drugs modulate/downregulate the high levels of TNFα seen during infection or induction? Would this allow domestic pigs to recover, rather than being rapidly killed by lymphopaenia and haemorrhagic shock? Could a small interfering RNA (siRNA) be directed against TNFα (or components of the TNF signalling pathway) to knock-down TNFα expression? Would this protect the animal from an excessive immune response? Could a transgenic animal (pig or mouse) be generated with an engineered/altered TNFα promoter and what phenotype would it display?

The *third* approach is to further develop the *in silico* analysis of the TNFα promoter sequence. By broadening the binding site analysis to include more suids (including

the entire phylogeny dataset) and other mammalian species, an extended model of the TNFα promoter could be established. This may enable a more confident prediction as to which candidate sites appear to be removed from the evolutionary norm, possibly indicative of specific positive selection, and thus warrant further investigation.

6.3 A238L and p54 Targeted/Mimicked Proteins

6.3.1 Experimental Approach

The six host proteins with which A238L and p54 interact (calcineurin Aβ, NFAT2, cyclophilin A, p65, IκBα, light chain dynein) were amplified by RT-PCR from mRNA isolated from an ASFV-susceptible domestic pig, resistant warthog and phenotypically unknown babirusa. These amplicons were cloned into the pGEM T-Easy vector (Promega) and sequencing reactions carried out using BigDye™ reagents (Applied Biosystems). Results files were processed and contigs assembled using the Staden Package (Staden *et al.* 1999; Staden *et al.* 2002). The insert within each clone was sequenced in both directions to give 100% overlap, ensuring accuracy of the read.

6.3.2 Comparative Sequencing – Targets Conserved between the Warthog and Domestic Pig

The predicted primary protein structure was identical between the ASFV-resistant warthog and ASFV-susceptible domestic pig for Light Chain Dynein (LCD), Cyclophilin A (CypA), IκBα and the regulatory domain of NFAT2. Given that the CypA and LCD protein sequence is fully conserved between test species suggests a conserved functionality of these genes with respect to their interaction with A238L between suid species. The sequence conservation of both IκBα and NFAT2 indicates that the ability of A238L to behave as an NFAT2- or IκBα-homologue is not influenced by any structural variation, which may affect competition with the native

proteins between pig species. However, it is worth noting that while protein sequence may remain highly or entirely conserved for all candidate sequences, variation in nucleotide sequence (coding and non-coding) can have a profound effect on transcription, RNA stability and processing. This may significantly alter the expression profile of the protein, regardless of any structural variation.

Calcineurin Aβ (CnAβ) nucleotide sequence is 1545bp in length in the domestic pig and babirusa and 1578bp long in the warthog. The most 5' ~118bp of all suid CnAβ sequences was not amplified. Compared to the domestic pig and babirusa, warthog CnAβ has a 3bp and a 30bp insertion; the 3bp insertion adds a single valine residue and the 30bp insertion adds ten residues. These amino acid insertions identified in warthog CnAβ are identical to a one amino acid and ten amino acid insertion found in two respective splice variants of human CnAβ, 'PPP3CB' and 'Q8N3W4' (http://www.ensembl.org/Homo_sapiens/). Both insertions occur outside the known functional domains of CnAβ. Furthermore, deletion analysis indicates that these splice events occur outwith the region necessary for A238L binding (Miskin *et al.* 2000). Therefore, it was considered that these differences in CnAβ are likely to be of little significance in determining ASF pathophysiology.

The babirusa was included in this study as phenotypically unknown with respect to ASFV pathophysiology. If the babirusa sequences had clearly resembled the warthog sequences, this may have been indicative of an evolutionary conserved ASFV resistance mechanism which the domestic pig has 'lost'. If instead the babirusa sequences had appeared more similar to the domestic pig, it could be postulated that the warthog has moved away from this conserved model in response to evolutionary pressure. However, this thesis clearly demonstrates that the domestic pig and warthog candidate sequences are predominantly more similar to each other than they are to the babirusa. This suggests that the sequence differences identified in the babirusa may be primarily due to species variation and evolutionary distance. Although these differences in the babirusa may be functional, their significance remains unclear at present.

6.3.3 Comparative Sequencing – p65

p65 coding sequence is 1662bp in length in the domestic pig and warthog and 1656bp in the babirusa and human. Three differences were identified between the predicted primary protein structure of warthog and domestic p65. Of particular interest is the presence of a proline (P) residue at position 531 in warthog p65, which is a conserved serine (S) residue in the other pigs and human. Indeed, this amino acid lies in the transactivation domain 1 (TA1) of p65 and has been well-characterised as an important casein kinase II phosphorylation site (ser^{529}) involved in secondary level upregulation of NFκB transcriptional activity in human p65 (Bird *et al.* 1997; Wang and Baldwin Jr. 1998; Wang *et al.* 2000). It was therefore hypothesised that the loss of this site from the warthog sequence may leave it less responsive to signal-induced activation. To confirm these sequencing results, a 268bp region encompassing all three differences was sequenced in a further fourteen warthogs, eight domestic pigs and six bushpigs (*Potamochoerus spp.*) (Figure 3.9). Like warthogs, bushpigs are also native to Africa and are resistant to African swine fever. The mutations at positions 448 and 485 were found to be conserved between warthogs and bushpigs; however the presence of a proline residue at position 531 is only seen in warthogs. The discovery of three differences in p65 has enabled the commercial sponsor of this work, Sygen International, to screen representatives of six domestic pig breeds (N=289) and thirteen synthetic lines currently used in commercial swine production worldwide (N=450) for the 'warthog-like' p65 allele (Table 6.1). However, to date their results indicate that the p65 allele may be fixed in all *Sus scrofa*.

Interestingly, the nucleotide difference which encodes a proline residue at position 531 in warthogs also introduces a unique *Sma* I restriction site into the 268bp amplicon (CCCGGG). In the domestic pig and babirusa (serine 531) this sequence is CTCGGG. This represents a potentially useful tool to screen large numbers of samples for this allele. PCR can be carried out directly from a DNA sample followed by restriction digest of the product, without the need to sequence each sample

Table 6.1

The commercial sponsor of this project, Sygen International, screened representatives of six domestic pig breeds (N=289) and thirteen synthetic lines currently used in commercial swine production worldwide (N=450) for the 'warthog-like' p65 allele. However, their results indicate that the p65 allele is fixed in all *Sus scrofa*.

Breed	Number of animals genotyped for p65
Landrace	(N=47)
Large White	(N-47)
Duroc	(N=47)
Hampshire	(N=47)
Pietrain	(N=94)
Meishan	(N=7)
Representatives of 13 synthetic lines	(N=450)

individually. Furthermore, false positives are unlikely to occur as this process aims to detect the presence of a restriction site, rather than its disappearance.

All three differences identified in p65 occur outside the rel homology domain (RHD), which is responsible for interaction with IκBα. It is therefore unlikely that these differences impact directly on physical interaction with either IκBα or A238L. Instead, they may cause indirect changes in NFκB activity, for example by altering interaction with transcriptional cofactors or histone-modifying enzymes (Campbell *et al.* 2004). Although six single nucleotide polymorphisms (SNPs) have been identified in the coding sequence of human p65 (Table 6.2), none have been described at serines 276, 529, and 536, which have been identified as sites of signal-induced phosphorylation (Wang and Baldwin Jr. 1998; Zhong *et al.* 1998; Sakurai *et al.* 1999). Therefore, the mutation from serine to proline at position 531 in the warthog sequence, equivalent to position 529 in human p65, may be of functional significance.

Despite its central role in the immune and inflammatory response and the existence of a large number of patent-protected human sequences in the EMBL database, few other mammalian species have been sequenced for p65. The only other mammals to have p65 sequences available in the public database are the mouse, rat and chimpanzee. It is also remarkable that despite current interest in autoimmunity and potential therapy offered by pig xenotransplantation, the suids represented within this study appear to be the only examples sequenced for p65. This lack of mammalian sequence data makes a comparative analysis of these sites somewhat limited. However, the three identified phosphorylation sites equivalent to human serines 276, 529 and 536 are conserved between the mammalian species available, except for the proline at position 531 (human 529) in the warthog (Figure 6.1).

NFκB and p65 are components of a highly evolutionarily conserved system which is essential to both immunity and development. It is therefore surprising that sequence variation has been identified in a region previously demonstrated to be of functional importance. However, acute ASFV infection is associated with a massive, largely

Table 6.2

The six single nucleotide polymorphisms (SNPs) have been identified in human p65 (Ensembl genome browser). However, none of these occur at the previously identified inducible-phosphorylation sites or in the sites which are equivalent to those differing between the suid species.

Residue	SNP ID	SNP type	Alleles	Ambiguity code	Alternate residues
96	11557248	Non-synonymous	C/A	M	R, L
197	11606329	Non-synonymous	A/G	R	C, R
313	12721572	Non-synonymous	T/C	Y	M, V
377	7116571	Synonymous	C/T	Y	-
381	12721574	Synonymous	T/C	Y	-
442	11608099	Non-synonymous	A/C	M	H, Q

Figure 6.1 – Conservation across Mammalian p65

SS	*Sus scrofa*	Domestic Pig
PA	*Phacochoerus africanus*	Warthog
BB	*Babyrousa babyrussa*	Babirusa
HS	*Homo sapiens*	Human
PT	*Pan troglodytes*	Chimpanzee
MM	*Mus musculus*	Mouse
RN	*Rattus norvegicus*	Rat

Comparative alignment of mammalian p65 protein sequences. The three inducible phosphorylation sites identified in human p65 (serines 276, 529 and 536) have been highlighted (numbering has been altered due to longer mouse sequence). All are conserved except for the proline at position 531 in warthog p65 (position 567 in this alignment).

```
SS ---------- ---------- ---------- --------DL FPLIFPSEPA PASGPYVEII EQPKQRGMRF  [ 70]
PA ---------- ---------- ---------- --------.. .......... .......... ..........  [ 70]
BB ---------- ---------- ---------- --------.. .......... .......... ..........  [ 70]
HS ---------- ---------- ---------- ------LSE. ......A... Q......... ..........  [ 70]
PT ---------- ---------- ---------- ------LSE. ......A... Q......... ..........  [ 70]
MM DATTGARRAF PPLANGFTLA RRGLSCDPGP APGTLTMD.. .......... Q......... ..........  [ 70]
RN ---------- ---------- ---------- ------LS.. .......... Q......... ..........  [ 70]

SS RYKCEGRSAG SIPGERSTDT TKTHPTIKIN GYTGPGTVRI SLVTKDPPHR PHPHELVGKD CRDGFYEAEL  [140]
PA .......... .......... .......... .......... .......... .......... ..........  [140]
BB .......... .......... .......... .......... .......... .......... ..........  [140]
HS .......... .......... .......... .......... .......... .......... ..........  [140]
PT .......... .......... .......... .......... .......... .......... ..........  [140]
MM .......... .......... .......... .......... .......... .......... ....Y...D.  [140]
RN .......... .......... .......... .......... .......... .......... ..........  [140]

SS CPDRCIHSFQ NLGIQCVKKR DLEQAINQRI QTNNNPFQVP IEEQRGDYDL NAVRLCFQVT VRDPAGRPLR  [210]
PA .......... .......... .......... .......... .......... .......... ..........  [210]
BB .......... .......... ......S... .......... .......... .......... ..........  [210]
HS .......... .......... ......S... .......... .......... .......... ....S.....  [210]
PT .......... .......... ......S... .......... .......... .......... ..E.S.....  [210]
MM ....S..... .......... ......S... .......H.. .......... .......... .........L  [210]
RN .......... .......... ......S... .......... .......... .......... ....S.....  [210]

SS LPPVLSHPIF DNRAPNTAEL KICRVNRNSG SCLGGDEIFL LCDKVQKEDI EVYFTGPGWE ARGSFSQADV  [280]
PA .......... .......... .......... .......... .......... .......... ..........  [280]
BB .......... .......... .......... .......... ......E... .......... ..........  [280]
HS .......... .......... .......... .......... .......... .......... ..........  [280]
PT .......... .......... .......... .......... .......... .......... ..........  [280]
MM .T........ .......... .......... .......... .......... .......... ..........  [280]
RN .T........ .......... .......... .......... .......... .......... ..........  [280]

SS HRQVAIVFRT PPYADPSLQA PVRVSMQLRR SDRELSEPM EFQYLPDTDD RHRIEEKRKR TYETFKSIMK  [350]
PA .......... .......... .......... .......... .......... .......... ..........  [350]
BB .......... ...S...... .......... .......... .......... .......... ..........  [350]
HS .......... .......... .......... .......... .......... .......... ..........  [350]
PT .......... .......... .......... .......... .......... .......... ..........  [350]
MM .......... .......... .......... .......... .......... .......... ..........  [350]
RN .......... .......... .......... .......... .......... .......... ..........  [350]

SS KSPFNGPTDP RPATRRIAVP SRSSASVPKP APQPYPFTPS LSTINFDEFT PMAFASGQIP GQTSALAPAP  [420]
PA .......... .......... .......... .......... .......... .......... ..........  [420]
BB .......... .......... .......... .......... .......... .......... ...A..V...  [420]
HS ....S..... ..PP...... .......... ........S. .....Y...P T.V.P....S -.A.......  [420]
PT ....S..... ..PP...... .......... ........S. .....Y...P T.V.P....S -.A.......  [420]
MM ........E. ..P....... T.N.T..... .....T.PA. .........S ..LLP....S N.AL....S-  [420]
RN ........E. ..PP...... ..GPT..... .....A.ST. .........S ..VLPP...S N.AL....S-  [420]

SS APVLVQAPAP APAPAMASAL AQAPAPVPVL APGLAQAVAP PAPKTNQAGE GTLTEALLQL QFDTDEDLGA  [490]
PA .......... .......... .......... .......... .......... .......... ...A......  [490]
BB .......... ...--..... .......... .......... .......... .......... ...A......  [490]
HS PQ..P..... ......V... .......... ...PP..... ....PT.... ...S...... ...-......  [490]
PT PQ..P..... ......V... .......... ...PP..... ....PT.... ...S...... ...-......  [490]
MM -SAP.L.QTM V.SS..VP-. ..P...A... T..PP.SLSA .V..ST.... ...S....H. ...A......  [490]
RN -SAP.L.QTM V.SS..VPS. ..P....... ...PP.SLSA .V..ST.... ...S....H. ...A......  [490]

SS LLGNNTDPTV FTDLASVDNS EFQQLLNQGV SMPPHTAEPM LMEYPEAITR LVTGSQRPPD PAPTPLGASG  [560]
PA .......... .......... .......... P......... .......... .......... ..........  [560]
BB .......... .......... .S........ P......... .......... .......... ..........  [560]
HS ....S...A. .......... .........I PVA...T... .......... ....A..... ...A....P.  [560]
PT ....S...A. .......... .........I PVA...T... .......... ....A..... ...A....P.  [560]
MM ....S...G. .......... .......... ..SHS..... .......... .......... .......T..  [560]
RN ........G. .......... .......... A.SHS..... .......... .......... ...AT..T..  [560]

SS LTNGLLSGDE DFSSIADMDF SALLSQISS [589]
PA ......P... .......... ......... [589]
BB .......... .......... ......... [589]
HS .P........ .......... ......... [589]
PT .P........ .......... ......... [589]
MM .P.-G..... .......... ......... [589]
RN .P.-G..... .......... ......... [589]
```

NFκB-mediated, release of proinflammatory cytokines (Gómez del Moral *et al.* 1999; Carrasco *et al.* 2002; Salguero *et al.* 2002). It is conceivable that any tempering of NFκB activity in the warthog, for example by removing such a phosphorylation site, may help to prevent the 'cytokine storm' seen in response to acute ASFV infection in the domestic pig.

Although only limited work has been undertaken on ASFV infection of the warthog and bushpig, they appear to have evolved different mechanisms of resistance to ASFV (Thomson *et al.* 1980; Anderson *et al.* 1998; Oura *et al.* 1998a). The length of viraemia and modes of transmission are clearly different between species (see section 1.7), despite the recently constructed molecular phylogeny indicating that they are monophyletic in origin (Figure 3.13) (Lowden *et al.* 2004). This may suggest that the common ancestor of these species, which was initially isolated in sub-Saharan Africa, might have diverged before being exposed to ASFV. Alternatively, isolated populations of the same ancestral species may have developed resistance by independent mechanisms before later becoming separate species. In terms of sequence differences in p65, the fact that the warthog has a potentially important difference, which is not present in the bushpig, could reflect their independently acquired ASFV resistance mechanisms.

Although the position of p65 has never been mapped in the domestic pig, it would appear to be located on the p arm of chromosome two. This prediction is based on the high conservation of synteny which exists between p arm and partial q arm of human chromosome eleven and the p arm of domestic pig chromosome two (cytogenetic map of the pig: http://www.toulouse.inra.fr/lgc/pig/cyto/cyto.htm). This prediction is further strengthened by the location of two genes which flank the human p65 gene that have been mapped to this region in both species. The calpain 1 gene (CAPN1) is located 0.5 centimorgans (cM) centromeric and the cathepsin F precursor (CTSP) gene lies 1cM telomeric to the p65 (RELA) gene. Given the immunobiological importance of p65 and the potential significance of sequence variation identified in this study, in collaboration with Professor Alan Archibald (Roslin Institute), it is intended that the position of p65 be precisely mapped in the domestic pig.

6.3.4 Comparative Activity – p65

Domestic pig and warthog p65-expression plasmids have been created and transfected into Vero cells together with an NFκB-luciferase reporter plasmid to investigate functionality attached to differences identified between these species (Chapter 5). This pilot study in Vero cells indicates that the presence of warthog p65 increased the transcriptional activity of NFκB significantly more than domestic pig p65 in response to LPS and TNFα ($p<0.01$). Transiently expressed A238L appeared to significantly sensitise cells expressing either domestic pig or warthog p65 to induction ($p<0.01$).

6.3.4.1 Experimental Considerations

As was the case for the TNFα promoter experiments, it is important to recognise that in all *in vitro* experiments involving tissue culture there are many factors which remove it from the *in vivo* situation. The objective was to create a controlled, albeit experimental, system to investigate a hypothesis, not to perfectly mirror the *in vivo* environment. However, several key questions regarding the suitability and validity of this system should be addressed. A key consideration is whether the reporter system (in this case a synthetic κB promoter) provides a genuine impression of NFκB activity in response to the different p65 genotypes. The behaviour of the synthetic κB promoter is liable to be have been influenced by its chromatin organisation and histone structure, which is unlikely to resemble that of an *in vivo* κB response element. Furthermore, if sequence differences in p65 affect phosphorylation and/or association with cofactors as hypothesised, do these cofactors have the opportunity to express their activity on the synthetic promoter? As previously described, many transcription factors bind cooperatively at composite elements. This basic 4xκB promoter may not be able to interact with and respond to active transcription factor complexes. The precise sequence of the κB sites in the synthetic promoter is also an issue. Several positions within the 10bp κB consensus site are variable and may

determine specificity for particular NFκB dimers. It has also recently been demonstrated that sequence variation in the binding site influences the conformation of the bound transcription factor. This in turn dictates whether or not the transcription factor forms productive interactions with coregulators, thus affecting transcriptional activity (Leung *et al.* 2004).

In order to address some of the issues described above, a promoter with 'real' NFκB response elements could be used to drive expression of the luciferase reporter protein. Both partial (the proximal 318bp) interleukin 8 promoter (Powell *et al.* 1996) and the IκBα promoter (de Martin *et al.* 1993) have been used previously for this purpose. However, such sequences contain many other binding sites and/or regulatory domains and introduce a range of unknown factors. Furthermore, even if one of these 'genuine' promoters was used, when present as a transgene its chromatin structure, chromosomal context and thus expression profile may be significantly altered.

As discussed regarding the TNFα promoter experiment, there is also the question of cell type and species. Indeed, the comparative analysis described in Chapter 3 clearly demonstrates that there are numerous sequence differences between p65 in suids and primates. This raises the question of species compatibility between the transcription machinery in the Vero (monkey) cells and the suid p65 and synthetic κB transgenes. However, we are again left with the predicament of using a porcine (domestic pig) line versus biasing our results against the warthog protein. In this experiment, cell type is not such an important issue as NFκB is present and functions in most cell types in response to cytokine-mediated and other immunogenic stimuli.

Although all the above issues draw attention to the gap which exists between *in vitro* and *in vivo* studies, as previously stressed, the aim was to devise a system which offered a method for comparing two p65 genotypes. This has been achieved using a system which may actually enable a clearer interpretation of results by removing some of the complex levels of control and interactions which exist *in vivo*.

6.3.4.2 Interpretation of Results

During the analysis of comparative p65 activity (Section 5.2), the uninduced cells unexpectedly displayed higher relative levels of luciferase activity than some of the induced cells. This appears to be due to low total protein concentration in the lysate of uninduced cells. When total protein concentration was used to normalise the luciferase activity, the uninduced cells therefore displayed a higher relative activity. This difference in protein concentration could be due to the inducing agents (and/or the process of transfection) acting as mitogens. Such an affect would cause cellular activation and proliferation, leading to higher total protein concentrations in the induced/transfected cells. More likely however, given the relatively short time scale, is that induction and the transfection process may have affected the adhesion of the cells to the plate. For example, induced and transfected cells may have detached more readily from the plate during the *in situ* lysis stage, resulting in a greater yield of protein in the lysate (e.g. membrane proteins). In contrast, uninduced cells may have simply lysed, leaving a majority of their membranes and associated proteins attached to the culture plate.

From these comparative activity results, it appears that warthog p65 significantly increases the transcriptional activity of NFκB, compared to domestic pig p65, when induced with LPS and TNFα ($p<0.01$). These findings conflict with our hypothesis that the existence of a proline residue in place of serine at position 531 in the warthog protein would decrease its activity by removing an inducible phosphorylation site. It is therefore possible that the other two sequence differences within the transactivation domains may be acting in a compensatory manner. To investigate the individual roles played by each of the three sequence differences, each could be systematically mutated 'back' to the domestic pig allele. This could be achieved using systems such as the QuikChange Site-Directed Mutagenesis Kit (Stratagene). This procedure uses primer sequence and PCR to introduce a single site mutation into a plasmid. As mentioned previously, it is also possible that the synthetic κB promoter does not allow any differences in NFκB activity to be reflected in a true manner.

Despite being in contrast to the initial hypothesis, there are a number of reasons why the warthog may encode more transcriptionally active p65/NFκB. These differences, if a true representation, may simply reflect a compensatory mechanism for the low TNFα promoter activity demonstrated in warthogs in Chapter 4. Alternatively, as NFκB is a highly conserved pathway it may be targeted by a number of pathogens in addition to ASFV. Therefore, the warthog may express more active p65 to combat this effect.

The ASFV protein, A238L, has previously been described as an IκBα homologue and an effective NFκB repressor (Powell *et al.* 1996; Miskin *et al.* 1998; Miskin *et al.* 2000; Tait *et al.* 2000). The fact that A238L significantly increased the sensitivity of NFκB to induction in cells expressing both warthog and domestic pig p65 ($p<0.05$), was highly unexpected. Analysis of Variance (ANOVA) indicated that the effect mediated by A238L was independent of p65 genotype. This can be explained by the fact that the sequence differences in p65 are located outside the rel homology domain with which A238L is proposed to interact (Powell *et al.* 1996; Tait *et al.* 2000). One possible explanation for the observed action of A238L is that there was competition for the response elements in the synthetic κB promoter. This may have been between NFκB containing endogenous and exogenous p65 or between different forms of endogenous NFκB dimers. By adding A238L, one form of NFκB may have been selectively removed, altering the equilibrium/dynamics of this interaction in favour of a more transcriptionally active complex.

Comparative analysis of A238L in low and high virulence strains of ASFV demonstrates complete sequence conservation (Neilan *et al.* 1997; Gil *et al.* 2003). However, it has been reported that a low virulence strain expressed higher levels of A238L and induced higher expression of proinflammatory cytokines (including TNFα) than the highly virulent strain (Gil *et al.* 2003). A marked increase in TNFα expression by macrophages infected with virulent ASFV isolates has been demonstrated *in vitro* by both Gómez del Moral *et al.* (1999) and Carrasco *et al.* (2002). These findings were confirmed *in vivo* by Salguero *et al.* (2002). Furthermore, an A238L-knockout form of a virulent ASFV strain remains highly

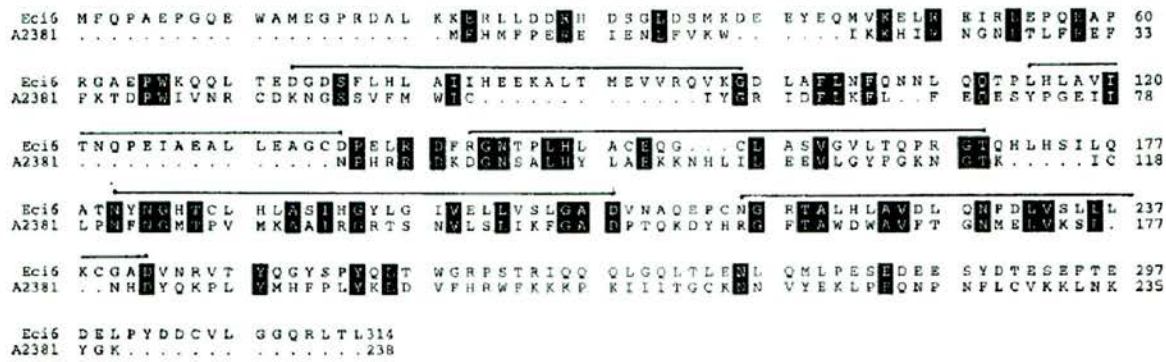
lethal to domestic pigs (Neilan *et al.* 1997). This is coupled with the fact that A238L lacks the first ankyrin repeat present in porcine IκBα (Figure 6.2), which is primarily responsible for strong inhibitory binding to NFκB subunits (Simeonidis *et al.* 1999). These *in vitro* and *in vivo* studies do not support an immunomodulatory function for A238L in domestic pigs. However, it is possible that this gene is involved in modulating the infection process in the highly-adapted, coevolved hosts (warthogs and/or the *Ornithodoros* tick). However, as previously stated, the ANOVA indicated that A238L functioned in the same manner, regardless of p65 genotype. This would suggest that if A238L does have a species-specific function, it does not appear to be related to polymorphisms in p65.

In contrast to the above reports, *in vitro* ASFV infection of macrophages has been shown to prevent PMA-induced proinflammatory cytokine expression (Powell *et al.* 1996). Furthermore, using an NFκB-responsive region of the IL-8 promoter, Powell *et al.* (1996) demonstrated that the NFκB-mediated transcription is inhibited by virus infection. This was specifically due to a 32kDa form of A238L which interacts with p65 (Tait *et al.* 2000). ASFV-infection of endothelial cells has also been shown to reduce proinflammatory cytokine production (IL-6 and IL-8) in response to TNFα stimulation (Vallée *et al.* 2001). Furthermore, Miskin *et al.* (1998) demonstrated interaction between the 28kDa form of A238L and components of the NFAT immunomodulatory pathway. Both A238L expression and ASFV infection consistently reduced NFAT-dependent reporter gene expression. This inhibitory effect was not seen when using an A238L-knockout strain of ASFV (Miskin *et al.* 1998; Miskin *et al.* 2000).

Thus it appears that current opinion on the function of A238L is somewhat divided. Work undertaken in IAH Pirbright strongly supports the hypothesis that A238L is involved in modulating the host immune response (Powell *et al.* 1996; Miskin *et al.* 1998; Miskin *et al.* 2000; Tait *et al.* 2000; Vallée *et al.* 2001). In contrast, research by Spanish and Portuguese groups demonstrates that ASFV infection induces massive cytokine production, with little evidence for A238L-mediated cytokine

Figure 6.2 – Sequence Comparison of IκBα and A238L

Sequence alignment of porcine IκBα (Eci6) and A238L taken from Powell *et al.* (1996). The five ankyrin repeats in of porcine IκBα have been marked by horizontal lines above the sequence. The conserved amino acids have been highlighted. A238L does not encode an ankyrin repeat homologous to the first repeat of porcine IκBα, which is considered to be primarily responsible for its strong inhibitory function (Simeonidis *et al.* 1999).



suppression (Gómez del Moral *et al.* 1999; Carrasco *et al.* 2002; Salguero *et al.* 2002; Gil *et al.* 2003).

If the A238L-mediated increase in luciferase activity presented in this thesis (Chapter 5) accurately reflects NFκB activity, which is in turn correlated to cytokine expression, these findings appear to be in disagreement with the *in vitro* studies which describe an A238L-driven decrease in proinflammatory cytokine expression (Powell *et al.* 1996; Tait *et al.* 2000). However, the luciferase expression attributable to the synthetic κB promoter may not be representative of the activity seen from a genuine proinflammatory cytokine promoter. Furthermore, NFκB activation can have both positive and negative regulatory effects, depending on the target promoter and inducing agent. It does not necessarily correlate with an increase in transcriptional activity (Campbell *et al.* 2004; Perkins 2004). Therefore, measuring cytokine expression may give a better impression of overall influence on the immune response. However, this process would involve many cellular factors, making any clear interpretation of the precise role of NFκB difficult.

6.3.5 Future p65 Work

This pilot study has raised many intriguing questions regarding the function of sequence differences in p65 and the role of A238L. Not only does the warthog p65 appear to be more transcriptionally active than that of the domestic pig, but A238L seems to sensitise NFκB to induction.

One obvious issue to be addressed first is the high levels of endogenous p65 detected by Western blot. This may be obscuring clear detection of differences in NFκB activity between the various experimental conditions. In order to remove this inference and increase resolution within the results, the three transfections could be repeated in a p65 null (-/-) cell line. Such a line has recently been donated to us by Professor Ron Hay (University of St. Andrews) and is derived from mouse embryonic fibroblasts (MEFs). To take this work forward, Sygen International is

supporting a PhD studentship from September 2004 in association with the Genesis Faraday Partnership.

Initial analysis of the p65 -/- MEFs has yielded an unexpected finding; they are remarkably tolerant to G418-mediated toxicity. Despite reports that cells with compromised NFκB activity are acutely sensitive to TNFα-mediated apoptosis (Zhou *et al.* 2003), a concentration of 4-6mg/ml G418 was required to kill p65 -/- MEFs, compared to 200μg/ml for the wild type MEFs (p65 +/+) (assay kindly undertaken by Sharon Watters, Roslin Institute). Although G418 and TNFα operate via separate mechanisms, this striking phenotypic difference between p65 -/- and wild type cells is worth considering during any subsequent analysis. However, it is possible that these cells carry G418-resistance from a previous iatrogenic source.

Once stably transfected p65 -/- cells have been generated, perhaps using an alternative selectable marker such as puromycin, the induction experiments and Western blot for p65 may be repeated. If necessary, exogenous p65 could first be separated from all other proteins by immunoprecipitation. The phosphorylation status of p65 could also be determined by Western blot using an anti-phosphoserine antibody or P^{32} (Wang *et al.* 2000). This approach may indicate if the serine/proline difference at position 531 affects inducible phosphorylation.

A number of additional factors may also be further investigated. These include using additional inducing agents, a range of inducing agent concentrations and viral infection. As mentioned previously, the individual roles played by the three sequence differences between warthog and domestic pig p65 could be dissected by mutational analysis. In addition, it may be possible to directly measure cytokine levels, for example using an ELISA system or real time PCR. This may give a more accurate impression of differences in activity caused by the p65 genotype, avoiding the use of a synthetic promoter. Alternatively, the use of promoter sequence containing a 'real' NFκB response element could be investigated. However, if the luciferase system was to be used, it may be best to first create a 'standard' cell line containing the reporter

plasmid, into which the experimental p65-expression vectors could then be subsequently transfected.

6.4 Conclusions

Comparative sequence analysis of six proteins targeted or mimicked by ASFV has been undertaken in the susceptible domestic pig (*Sus scrofa*), resistant warthog (*Phacochoerus africanus*) and phenotypically unknown babirusa (*Babyrousa babyrussa*). The extent and degree of sequence variation between these species has been assessed to identify gene regions which may be capable of modifying interaction with the viral proteins A238L and p54. These differences may determine the different pathophysiological outcomes of ASFV infection in suid species of varying susceptibility. Despite high levels of conservation in all candidate sequences (light chain dynein, cyclophilin A, NFAT2, IκBα, calcineurin Aβ and p65), three amino acid differences have been identified between domestic pig and warthog p65. As Sygen International is considering applying for patent protection of all p65 sequencing data, this work has not yet been published.

In addition, the TNFα promoter has been sequenced in the domestic pig, warthog and babirusa. Thirty nucleotide differences and three deletion events were identified between the domestic pig and warthog. These differences define the existence of putative transcription factor binding sites and modules, which may contribute to the cytokine-mediated pathological outcome of ASFV infection. All TNFα promoter sequence data presented within this thesis is first being submitted as part of a molecular phylogeny study, currently in preparation for publication. It is therefore intended to publish these data once the phylogeny paper has been accepted.

Against this background of consistent sequence conservation, the TNFα promoter and p65 studies were extended to investigate potential functionality attached to the identified sequence differences. This was achieved by comparing the activity of domestic pig and warthog genotypes by the use of luciferase reporter constructs *in vitro*. A novel dual-luciferase reporter system was developed to compare the two

TNFα promoters. This study indicates that the domestic pig TNFα promoter is significantly more transcriptionally active than that of the warthog. This was the case both when uninduced and induced using LPS and TNFα. Furthermore, the domestic pig TNFα promoter responds significantly more strongly to LPS than the warthog. These differences may reflect the surge of proinflammatory cytokine expression which occurs in domestic pigs following infection with ASFV.

A pilot study has been undertaken to investigate any difference in activity between warthog and domestic pig p65. Each p65 genotype was expressed to a high level in tissue culture and the activity of NFκB assessed using a luciferase reporter plasmid containing a synthetic 4xκB promoter. Contrary to expectations, these results indicate that NFκB was significantly more active in the presence of warthog p65 than domestic pig p65. Furthermore, co-expression of the ASFV protein A238L appeared to sensitise NFκB to induction, despite its previously proposed role as an NFκB repressor.

It is clear from these analyses that sequence differences in p65 and the TNFα promoter appear to have a significant effect on function *in vitro*. As discussed in depth, these results highlight an important need for further investigation. Variation in cytokine expression, whether mediated by a promoter or transcription factor, could have major implications for the immune and inflammatory response. Assuming that these *in vitro* findings mirror an *in vivo* phenomenon, p65 and TNFα promoter may offer key targets for therapeutic intervention in the form of drugs or genetic manipulation.

Although breeding programs could possibly be employed to introduce warthog alleles into domestic pigs, ethical issues and fertility/compatibility problems between these resistant and susceptible species suggest that this method would not be viable. Such an approach would also have serious repercussions for the genetic composition of commercial pig lines. In addition, given the results of the initial screen undertaken by Sygen International, it is highly improbable that a naturally occurring 'warthog-like' allele for p65 would be found in a domestic pig line. Together, these factors

suggest that transgenic strategies may represent the only feasible mechanism of introducing a particular genotype (whether synthetic or naturally occurring) into the commercial pig population. Such an approach could involve the use of exciting new technologies to generate pigs transgenic for functional, as opposed to purely experimental, purposes. Although this raises a plethora of ethical, philosophical and practical issues, disease resistance is just one possible application for the now-tangible development of future transgenic livestock.

7 References

Abu-Amer, Y, Ross, FP, Edwards, J and Teitelbaum, SL (1997). Lipopolysaccharide-stimulated Osteoclastogenesis Is Mediated by Tumor Necrosis Factor Via its P55 Receptor. *Journal of Clinical Investigation.* **100:**1557-1565.

Afonso, CL, Neilan, JG, Kutish, GF and Rock, DL (1996). An African swine fever virus Bcl-2 homolog, 5-HL, suppresses apoptotic cell death. *Journal of Virology.* **70:**4858-4863.

Aggarwal, BB (2003). Signalling pathways of the TNF superfamily: a double-edged sword. *Nature Reviews - Immunology.* **3:**745-756.

Aggarwal, BB, Kohr, WJ, Hass, PE, Moffat, B, Spencer, SA, Henzel, WJ, Bringman, TS, Nedwin, GE, Goeddel, DV and Harkins, RN (1985). Human Tumor Necrosis Factor. *Journal of Biological Chemistry.* **260:**2345-2354.

Alcami, A, Symons, JA and Smith, GL (2000). The Vaccinia Virus Soluble Alpha/Beta Interferon (IFN) Receptor Binds to the Cell Surface and Protects Cells from the Antiviral Effects of IFN. *Journal of Virology.* **74:**11230-11239.

Alonso, C, Miskin, J, Hernáez, B, Fernandez-Zapatero, P, Soto, L, Cantó, C, Rodríguez-Crespo, I, Dixon, L and Escribano, JM (2001). African Swine Fever Virus Protein p54 Interacts with the Microtubular Motor Complex through Direct Binding to Light-Chain Dynein. *Journal of Virology.* **75:**9819-9827.

Anderson, EC, Hutchings, GH, Mukarati, N and Wilkinson, PJ (1998). African swine fever infection of the bushpig (*Potamochoerus porcus*) and its significance to the epidemiology of the disease. *Veterinary Microbiology.* **62:**1-15.

Andres, GR, Garcia-Escudero, R, Simon-Mateo, C and Vinuela, E (1998). African swine fever virus is enveloped by a two-membraned collapsed cisterna derived from the endoplasmic reticulum. *Journal of Virology.* **72:**8988-9001.

Anon (2000). CSF: a communication crisis? New Agriculturist Online. http://www.new-agri.co.uk/00-6/focuson/focuson6.html.

Anon (2001). Recognizing African Swine Fever - A field manual. FAO Animal Health Manual No. 9. Rome, Food and Agriculture Organization of the United Nations.

Aronson, JF, Herzog, NK and Jerrells, TR (1995). Tumor Necrosis Factor and the Pathogenesis of Pichinde Virus Infection in Guinea Pigs. *American Journal of Tropical Medicine and Hygiene.* **52:**262-269.

Azeredo, EL, Zagne, SMO, Santiago, MA, Gouvea, AS, Santana, AA, Neves-Souza, PCF, Nogueira, RMR, Miagostovich, MP and Kubelka, CF (2001). Characterisation of Lymphocyte Response and Cytokine Patterns in Patients with Dengue Fever. *Immunobiology*. **204:**494-507.

Bäckhed, F and Hornef, M (2003). Toll-like receptor 4-mediated signalling by epithelial surfaces: necessity or threat? *Microbes and Infection*. **5:**951-9.

Baize, S, Leroy, EM, Georges, AJ, Georges-Courbot, M-C, Capron, M, Bedjabaga, I, Lansoud-Soukate, J and Mavoungou, E (2002). Inflammatory responses in Ebola virus-infected patients. *Clinical and Experimental Immunology*. **128:**163-168.

Barbar, E, Kleinman, B, Imhoff, D, Li, M, Hays, TS and Hare, M (2001). Dimerization and Folding of LC8, a Highly Conserved Light Chain of Cytoplasmic Dynein. *Biochemistry*. **40:**1596-1605.

Basta, S, Knoetig, S, Summerfield, A and McCullough, KC (2001). Lipopolysaccharide and phorbol 12-myristate 13-acetate both impair monocyte differentiation, relating cellular function to virus susceptibility. *Immunology*. **103:**488-497.

Baughman, RP, Lower, EE and du Bois, RM (2003). Sarcoidosis. *Lancet*. **361:**1111-1118.

Beato, M and Eisfeld, K (1997). Transcription factor access to chromatin. *Nucleic Acids Research*. **25:**3559-3563.

Beg, AA, Sha, WC, Bronson, RT and Baltimore, D (1995a). Constitutive NF-κB activation, enhanced granulopoiesis, and neonatal lethality in IκBα-deficient mice. *Genes & Development*. **9:**2736-2746.

Beg, AA, Sha, WC, Bronson, RT, Ghosh, S and Baltimore, D (1995b). Embryonic lethality and liver degeneration in mice lacking the RelA component of NF-κB. *Nature*. **376:**167-170.

Benedict, CA and Ware, CF (2001). Virus Targeting of the Tumor Necrosis Factor Superfamily. *Virology*. **289:**1-5.

Berger, SL (2002). Histone modification in transcriptional regulation. *Current Opinion in Genetics and Development*. **12:**142-148.

Bird, TA, Schooley, K, Dower, SK, Hagen, H and Virca, GD (1997). Activation of Nuclear Transcription Factor NF-κB by Interleukin-1 Is Accompanied by Casein Kinase II-mediated Phosphorylation of the p65 Subunit. *Journal of Biological Chemistry*. **272:**32606-32612.

Blom, A, Alers, MPT and Barnes, RFW (1990). Gabon. *In:* Antelopes - Global Survey and Regional Action Plans: Part 3. West and Central Africa. East, R (Ed.). Gland, IUCN. pp 171.

Borca, MV, Carrilllo, C, Zsak, L, Laegreid, WW, Kutish, GF, Neilan, JG, Burrage, TG and Rock, DL (1998). Deletion of a CD2-Like Gene, *8-DR*, from African Swine Fever Virus Affects Viral Infection in Domestic Swine. *Journal of Virology*. **72:**2881-2889.

Boreham, PFL and Geigy, R (1976). Culverts and Trypanosome Transmission in the Serengeti: III. Studies on the Genus *Auchmeromyia* Brauer and Bergenstamm (Diptera: Calliphoridae). *Acta tropica*. **33:**74-87.

Bosma, AA (1980). The karyotype of the babirusa (*Babyrousa babyrussa*); karyotype evolution of the Suidae. *14th European Colloquium of Cytogenetics in Domestic Animals*:238-241.

Bosma, AA and de Haan, NA (1981). The Karyotype of *Babyrousa babyrussa* (Suidae, Mammalia). *Acta Zoologica et Pathologica Antverpiensia*. **76:**17-27.

Bosma, AA, de Haan, NA and Macdonald, AA (1991). The current status of cytogenetics of the Suidae: a review. *Bongo, Berlin*. **18:**258-272.

Bosma, AA, de Haan, NA, Mellink, CHM, Yerle, M and Zijlstra, C (1996). Chromosome homology between the domestic pig and the babirusa (family Suidae) elucidated with the use of porcine painting probes. *Cytogenetics and Cell Genetics*. **75:**32-35.

Breese, SS and DeBoer, CJ (1966). Electron Microscope Observations of African Swine Fever Virus in Tissue Culture Cells. *Virology*. **28:**420-428.

Brun, A, Rivas, C, Esteban, M, Escribano, JM and Alonso, C (1996). African Swine Fever Virus Gene *A179L*, a Viral Homologue of *bcl-2*, Protects Cells from Programmed Cell Death. *Virology*. **225:**227-230.

Brun, A, Rodriguez, F, Escribano, JM and Alonso, C (1998). Functionality and cell anchorage dependence of the African swine fever virus gene A179L, a viral bcl-2 homolog, in insect cells. *Journal of Virology*. **72(12):**10227-10233.

Burkly, L, Hession, C, Ogata, L, Reilly, C, Marconi, LA, Olson, D, Tizard, R, Cate, R and Lo, D (1995). Expression of *relB* is required for the development of thymic medulla and dendritic cells. *Nature*. **373:**531-536.

Campbell, KJ, Rocha, S and Perkins, ND (2004). Active Repression of Antiapoptotic Gene Expression by RelA (p65) NF-κB. *Molecular Cell*. **13:**853-865.

Carrasco, L, Gómez-Villamandos, JC, Bautista, MJ, Martín de las Mulas, J, Villeda, CJ, Wilkinson, PJ and Sierra, AM (1996). In vivo replication of African swine fever virus (Malawi '83) in neutrophils. *Veterinary Research*. **27:**55-62.

Carrasco, L, Núñez, A, Salguero, FJ, Díaz San Segundo, F, Sánchez-Cordón, P, Gómez-Villamandos, JC and Sierra, MA (2002). African Swine Fever: Expression of Interleukin-1 alpha and Tumor Necrosis Factor-alpha by Pulmonary Intravascular Macrophages. *Journal of Comparative Pathology*. **126:**194-201.

Carrascosa, JL, Carazo, JM, Carrascosa, AL, García, N, Santisteban, A and Viñuela, E (1984). General Morphology and Capsid Fine Structure of African Swine Fever Virus Particles. *Virology*. **132:**160-172.

Carrilllo, C, Borca, MV, Afonso, CL, Onisk, DV and Rock, DL (1994). Long-Term Persistent Infection of Swine Monocytes/Macrophages with African Swine Fever Virus. *Journal of Virology*. **68:**580-583.

Carswell, EA, Old, LJ, Kassel, RL, Green, S, Fiore, N and Williamson, B (1975). An endotoxin-induced serum factor that causes necrosis of tumors. *Proceedings of the National Academy of Sciences of the United States*. **72:**3666-3670.

Carvalho, ZG, Alves De Matos, AP and Rodrigues-Pousada, C (1988). Association of African swine fever virus with the cytoskeleton. *Virus Research*. **11:**175-192.

Carver, TJ and Mullan, LJ (2002). A new graphical user interface to EMBOSS (http://www.hgmp.mrc.ac.uk/Software/EMBOSS/Jemboss/). *Comparative and Functional Genomics*. **3:**75-78.

Casal, I, Enjuanes, L and Viñuela, E (1984). Porcine Leukocyte Cellular Subsets Sensitive to African Swine Fever Virus In Vitro. *Journal of Virology*. **52(1):**37-46.

Chen, FE, Huang, D-B, Chen, Y-Q and Ghosh, G (1998a). Crystal structure of p50/p65 heterodimer of transcription factor NF-κB bound to DNA. *Nature*. **391:**410-413.

Chen, L, Glover, JNM, Hogan, PG, Rao, A and Harrison, SC (1998b). Structure of the DNA-binding domains from NFAT, Fos and Jun bound specifically to DNA. *Nature*. **392:**42-48.

Cheng, EH-Y, Nicholas, J, Bellows, DS, Hayward, GS, Guo, H-G, Reitz, MS and Hardwick, JM (1997). A Bcl-2 homolog encoded by Karposi sarcoma-associated virus, human herpesvirus 8, inhibits apoptosis but does not heterodimerize with Bax or Bak. *Proceedings of the National Academy of Sciences of the United States*. **94:**690-694.

Cheng, Q, Cant, CA, Moll, T, Hofer-Warbinek, R, Wagner, E, Birnstiel, ML, Bach, FH and de Martin, R (1994). NF-κB Subunit-specific Regulation of the IκBα Promoter. *Journal of Biological Chemistry*. **269:**13551-13557.

Cobbold, C, Brookes, SM and Wileman, T (2000). Biochemical Requirements of Virus Wrapping by the Endoplasmic Reticulum: Involvement of ATP and Endoplasmic Reticulum Calcium Store during Envelopment of African Swine Fever Virus. *Journal of Virology*. **74:**2151-2160.

Cobbold, C, Windsor, M and Wileman, T (2001). A Virally Encoded Chaperone Specialized for Folding of the Major Capsid Protein of African Swine Fever Virus. *Journal of Virology*. **75:**7221-7229.

Colgrove, GS (1968). Immunofluorescence and Inclusion Bodies in Circulating Leukocytes of Pigs Infected with African Swine Fever Virus. *Bulletin of Epizootic Diseases in Africa*. **16:**341-343.

Colgrove, GS, Haelterman, EO and Coggins, L (1969). Pathogenesis of African Swine Fever in Young Pigs. *American Journal of Veterinary Research*. **30:**1343-1359.

Cox, BF (1963). African Swine Fever. *Bulletin of Epizootic Diseases in Africa*. **11:**147-148.

Crépieux, P, Kwon, H, Leclerc, N, Spencer, W, Richard, S, Lin, R and Hiscott, J (1997). IκBα Physically Interacts with a Cytoskeleton-Associated Protein through Its Signal Response Domain. *Molecular and Cellular Biology*. **17:**7375-7385.

De Kock, G, Robinson, EM and Keppel, JJG (1940). Swine Fever in South Africa. *Onderstepoort Journal of Veterinary Science and Animal Industry*. **14:**31-93.

de Martin, R, Vanhove, B, Cheng, Q, Hofer, E, Csizmadia, V, Winkler, H and Bach, FH (1993). Cytokine-inducible expression in endothelial cells of an IkBa-like gene is regulated by NFkB. *The EMBO Journal*. **12(7):**2773-2779.

Dempsey, PW, Doyle, SE, He, JQ and Cheng, G (2003). The signaling adaptors and pathways activated by TNF superfamily. *Cytokine & Growth Factor Reviews*. **14:**193-209.

DeTray, DE (1963). African Swine Fever. *Advances in Veterinary Science*. **8:**299-333.

Dixon, LK, Abrams, CC, Bowick, G, Goatley, LC, Kay-Jackson, PC, Chapman, D, Liverani, E, Nix, R, Silk, R and Zhang, F (2004). African swine fever virus proteins involved in evading host defence systems. *Veterinary Immunology and Immunopathology*. **In press**.

Dixon, LK, Costa, JV, M, EJ, Rock, DL, Venuela, E and Wilkinson, PJ (2000). Family *Asfarviridae*. *In:* Virus Taxonomy: Seventh Report of the International Committee on Taxonomy of Viruses. van Regenmortel, MHV, Fanquet, CM, Bishop, DHL *et al* (Eds.). San Diego, Academic Press. pp 159-165.

Edwards, JF and Dodds, WJ (1985). Platelet and fibrinogen kinetics in healthy and African swine fever-affected swine: [^{75}Se]Selenomethionine-labeling study. *American Journal of Veterinary Research*. **46:**181-184.

Edwards, JF, Dodds, WJ and Slauson, DO (1984). Coagulation changes in African swine fever virus infection. *American Journal of Veterinary Research*. **45:**2414-2420.

Edwards, JF, Dodds, WJ and Slauson, DO (1985). Megakaryocytic Infection and Thrombocytopenia in African Swine Fever. *Veterinary Pathology*. **22:**171-176.

El Hicheri, K, Gomez-Tejedor, C, Penrith, M-L, Davies, G, Douati, A, Edoukou, GJ and Wojciechowski, K (1998). L'épizootie de peste porcine africaine de 1996 en Côte d'Ivoire. *Revue Scientifique et Technique de l'Office International des Épizooties*. **17:**660-673.

Els, HJ and Pini, A (1977). Negative Staining of a Non-Haemadsorbing Strain of African Swine Fever Virus. *Onderstepoort Journal of Veterinary Research*. **44(1):**39-46.

Enjuanes, L, Cubero, I and Viñuela, E (1977). Sensitivity of Macrophages from Different Species to African Swine Fever (ASF) Virus. *Journal of General Virology*. **34:**455-463.

Falvo, JV, Uglialoro, AM, Brinkman, BMN, Merika, M, Parekh, BS, Tsai, EY, King, HC, Morielli, AD, Peralta, EG, Maniatis, T, Thanos, D and Goldfeld, AE (2000). Stimulus-Specific Assembly of Enhance Complexes on the Tumor Necrosis Factor Alpha Gene Promoter. *Molecular and Cellular Biology*. **20:**2239-2247.

Fan, H, Peck, OM, Tempel, GE, Halushka, PV and Cook, JA (2004). Toll-like receptor 4 couples GI protein signaling pathways regulate extracellular signal-regulated kinase phosphorylation and Ap-1 activation independent of NFkappaB activation. *Shock*. **22:**57-62.

Farmer, P (1994). Swine Aid. *In:* The Haiti Files - Decoding the Crisis. Ridgeway, J (Ed.). Washington DC, Essential Books. pp 130-133.

Feldmann, H, Bugany, H, Mahner, F, Klenk, H-D, Drenckhahn, D and Schnittler, H-J (1996). Filovirus-Induced Endothelial Leakage Triggered by Infected Monocytes/Macrophages. *Journal of Virology*. **70:**2208-2214.

Fenner, FJ, Gibbs, EPJ, Murphy, FA, Rott, R, Studdert, MJ and White, DO (1993). Iridoviridae and African Swine Fever Virus. *In:* Veterinary Virology. San Diego, Academic Press. pp 391-401.

Fernández, A, Perez, J, Carrasco, L, Bautista, MJ, Sanchez-Vizcaino, JM and Sierra, AM (1992a). Distribution of ASFV Antigens in Pig Tissues Experimentally Infected with Two Different Spanish Virus Isolates. *Journal of Veterinary Medicine Reihe B*. **39:**393-402.

Fernández, A, Perez, J, Carrasco, L, Sierra, MA, Sanchez-Vizcaino, JM and Jover, A (1992b). Detection of African swine fever viral antigens in paraffin-embedded tissues by the use of immunohistologic methods and polyclonal antibodies. *American Journal of Veterinary Research*. **53:**1462-1467.

Fernández, A, Perez, J, Martin de las Mulas, J, Carrasco, L, Dominguez, J and Sierra, MA (1992c). Localization of African Swine Fever Viral Antigen, Swine IgM, IgG and Clq in Lung and Liver Tissues of Experimentally Infected Pigs. *Journal of Comparative Pathology*. **107:**81-90.

Foo, SY and Nolan, GP (1999). NF-κB to the rescue. RELs, apoptosis and cellular transformation. *Trends in Genetics*. **15:**229-235.

Frech, K, Danescu-Mayer, J and Werner, T (1997). A novel method to develop highly specific models for regulatory unit detects a new LTR in GenBank which contains a functional promoter. *Journal of Molecular Biology*. **270:**674-687.

Furutani, Y, Kato, A, Yasue, H, Alexander, LJ, Beattie, CW and Hirose, S (1998). Evolution of the trappin multigene family in the Suidae. *Journal of Biochemistry*. **124:**491-502.

Geisbert, TW, Hensley, LE, Gibb, TR, Steele, KE, Jaax, NK and Jahrling, PB (2000). Apoptosis induced *in vitro* and *in vivo* during infection by Ebola and Marburg viruses. *Laboratory Investigation*. **80:**171-186.

Gerritsen, ME, Williams, AJ, Neish, AS, Moore, S, Shi, Y and Collins, T (1997). CREB-binding protein/p300 are transcriptional coactivators of p65. *Proceedings of the National Academy of Sciences of the United States*. **94:**2927-2932.

Ghosh, S, May, MJ and Kopp, EB (1998). NF-κB and Rel Proteins: Evolutionary Conserved Mediators of Immune Responses. *Annual Review of Immunology*. **16:**225-260.

Gibbs, EPJ and Butler, JF (1984). African swine fever - an assessment of risk for Florida. *Journal of the American Medical Association*. **184:**644-647.

Gil, S, Spagnuolo-Weaver, M, Canals, A, Sepúlveda, N, Oliveira, J, Aleixo, A, Allan, G, Leitão, A and Martins, CLV (2003). Expression at mRNA level of cytokines and A238L gene in porcine blood-derived macrophages infected *in vitro* with *African swine fever virus* (ASFV) isolates of different virulence. *Archives of Virology*. **148:**2077-2097.

Gilmore, TD (2004). NF-kB.org (Rel/NF-kB Transcription Factors). http://people.bu.edu/gilmore/nf-kb/index.html.

Goldman, G, Soffer, D, Heller, L, Aderka, D, Lahat, A and Klausner, JM (2001). Tumour Necrosis Factor Mediates Bacterial Translocation After Haemorrhagic Shock Endotoxaemia. *European Journal of Surgery*. **167:**299-304.

Gómez del Moral, M, Ortuño, E, Fernández-Zapatero, P, Alonso, F, Alonso, C, Ezquerra, A and Domínguez, J (1999). African Swine Fever Virus Infection Induces Tumor Necrosis Factor Alpha Production: Implications in Pathogenesis. *Journal of Virology*. **73(3):**2173-2180.

Gómez-Villamandos, JC, Bautista, MJ, Hervás, J, Carrasco, L, Chacón-M. de Lara, F, Pérez, J, Wilkinson, PJ and Sierra, AM (1996). Subcellular Changes in Platelets in Acute and Subacute African Swine Fever. *Journal of Comparative Pathology*. **115:**327-241.

Gómez-Villamandos, JC, Hervás, J, Méndez, A, Carrasco, L, Martín de las Mulas, J, Villeda, CJ, Wilkinson, PJ and Sierra, MA (1995a). Experimental African swine fever: apoptosis of lymphocytes and virus replication in other cells. *Journal of General Virology*. **76:**2399-2405.

Gómez-Villamandos, JC, Hervás, J, Méndez, A, Carrasco, L, Villeda, CJ, Wilkinson, PJ and Sierra, MA (1995b). Ultrastructural study of the renal tubular system in acute experimental African swine fever: virus replication in glomerular mesangial cells and in the collecting ducts. *Archives of Virology*. **140:**581-589.

Greig, A (1972). The Localization of African Swine Fever Virus in the Tick *Ornithodoros Moubata porcinus*. *Archiv für die gesamte Virusforschung*. **39:**240-247.

Greig, A and Plowright, W (1970). The excretion of two virulent strains of African swine fever virus by domestic pigs. *Journal of Hygiene*. **68:**673-682.

Groocock, CM, Hess, WR and Gladney, WJ (1980). Experimental Transmission of African Swine Fever Virus by *Ornithodoros coriaceus*, an Argasid Tick Indigenous to the United States. *American Journal of Veterinary Research*. **41:**591-594.

Groves, CP and Grubb, P (1993). The Eurasian Suids *Sus* and *Babyrousa*: Taxonomy and Description. *In:* Pigs, Peccaries and Hippos. Oliver, WLR (Ed.). Gland, IUCN. pp 107-111.

Grubb, P (1993). The Afrotropical Suids *Phacochoerus* and *Potamochoerus*: Taxonomy and Description. *In:* Pigs, Peccaries and Hippos. Oliver, WLR (Ed.). Gland, IUCN. pp 66-75.

Guo, Z, Boekhoudt, GH and Boss, JM (2003). Role of the intronic enhancer in tumor necrosis factor-mediated induction of manganous superoxide dismutase. *Journal of Biological Chemistry*. **278:**23570-23578.

Haig, DM (2001). Subversion and piracy: DNA viruses and immune evasion. *Research in Veterinary Science*. **70:**205-219.

Hajeer, AH and Hutchinson, IV (2000). TNF-α Gene Polymorphism: Clinical and Biological Implications. *Microscopy Research and Technique*. **50:**216-228.

Hajeer, AH and Hutchinson, IV (2001). Influence of TNFα Gene Polymorphisms on TNFα Production and Disease. *Human Immunology*. **62:**1191-1199.

Harada, A, Takei, Y, Kanai, Y, Tanaka, Y, Nonaka, S and Hirokawa, N (1998). Golgi Vesiculation and Lysosome Dispersion in Cells Lacking Cytoplasmic Dynein. *Journal of Cell Biology*. **141:**51-59.

Haudek, SB, Natmeßnig, BE, Redl, H, Schlag, G and Giroir, BP (1998). Genetic sequences and transcriptional regulation of the *TNFA* promoter: comparison of human and baboon. *Immunogenetics*. **48:**202-207.

Heath, CM, Windsor, M and Wileman, T (2001). Aggresomes Resemble Sites Specialized for Virus Assembly. *The Journal of Cell Biology*. **153:**449-455.

Henderson, S, Huen, D, Rowe, M, Dawson, C and Johnson, G (1993). Epstein-Barr virus-coded BHRF1 protein, a viral homologue of Bcl-2, protects human B cells from programmed cell death. *Proceedings of the National Academy of Sciences of the United States*. **90:**8479-8483.

Hess, WR (1981). African Swine Fever: A Reassessment. *Advances in Veterinary Science and Comparative Medicine*. **25:**39-69.

Hess, WR, Endris, RG, Haslett, TM, Monahan, MJ and McCoy, JP (1987). Potential Arthropod Vectors of African Swine Fever Virus in North America and the Caribbean Basin. *Veterinary Parasitology*. **26:**145-155.

Heuschele, WP and Coggins, L (1965a). Isolation of African Swine Fever Virus from a Giant Forest Hog. *Bulletin of Epizootic Diseases of Africa*. **13:**255-256.

Heuschele, WP and Coggins, L (1965b). Studies on the Transmission of African Swine Fever Virus by Arthropods. *In:* Sixty-ninth Annual Proceedings of the United States Livestock Sanitary Association. pp pp94-100.

Heuschele, WP and Coggins, L (1969). Epizootiology of African swine fever virus in warthogs. *Bulletin of Epizootic Diseases in Africa*. **17:**179-183.

Hober, D, Poli, L, Roblin, B, Gestas, P, Chungue, E, Granic, G, Imbert, P, Pecarere, J-L, Vergez-Pascal, R, Wattre, P and Maniez-Montreuil, M (1993). Serum levels of tumor necrosis factor-α (TNF-α), interleukin-6 (IL-6), and interleukin-1β (IL-1β) in Dengue-infected patients. *American Journal of Tropical Medicine and Hygiene*. **48:**324-331.

Hogan, PG, Chen, L, Nardone, J and Rao, A (2003). Transcriptional regulation by calcium, calcineurin, and NFAT. *Genes & Development*. **17:**2205-2232.

Holzbaur, ELF and Vallee, RB (1994). Dyneins: Molecular Structure and Cellular Function. *Annual Review of Cell Biology*. **10:**339-372.

Horst, HS, Dijkhuizen, AA, Huirne, RBM and De Leeuw, PW (1998). Introduction of contagious animal diseases into The Netherlands: elicitation of expert opinions. *Livestock Production Science*. **53:**253-264.

Horwitz, BH, Scott, ML, Cherry, SR, Bronson, RT and Baltimore, D (1997). Failure of lymphopoiesis after adoptive transfer of NF-κB-deficient fetal liver cells. *Immunity*. **6:**765-772.

Hu, Y, Baud, V, Delhase, M, Zhang, P, Deerinck, T, Ellisman, M, Johnson, R and Karin, M (1999). Abnormal morphogenesis but intact IKK activation in mice lacking the IKKα subunit of IκB kinase. *Science*. **284:**316-320.

Iizuka, M and Smith, MM (2003). Functional consequences of histone modifications. *Current Opinion in Genetics and Development*. **13:**154-160.

Irwin, DM, Kocher, TD and Wilson, AC (1991). Evolution of the Cytochrome *b* Gene of Mammals. *Journal of Molecular Evolution*. **32:**128-144.

Jaffrey, SR and Snyder, SH (1996). PIN: An Associated Protein Inhibitor of Neuronal Nitric Oxide Synthase. *Science*. **274:**774-777.

Jenuwein, T and Allis, CD (2001). Translating the Histone Code. *Science*. **293:**1074-1080.

Johnston, HH (1905). Notes on mammals and birds of Liberia. *Proceedings of the Zoological Society of London*:197-291.

Johnston, HH (1906). Liberia. London.

Johnston, HH (1908). George Grenfell and the Congo. London, Hutchinson.

Kaijzel, EL, Bayley, J-P, van Krugten, MV, Smith, L, van de Linde, P, Bakker, AM, Breedveld, FC, Huizinga, TWJ and Verweij, CL (2001). Allele-specific quantification of tumor necrosis factor α (TNF) transcription and the role of promoter polymorphisms in rheumatoid arthritis patients and healthy individuals. *Genes and Immunity*. **2:**135-144.

Karin, M (1999). The beginning of the end: IκB kinase (IKK) and NF-κB activation. *Journal of Biological Chemistry*. **274:**27339-27342.

Karin, M and Ben-Neriah, Y (2000). Phosphorylation Meet Ubiquitination: The Control of NF-κB Activity. *Annual Review of Immunology*. **18:**621-663.

Kassiotis, G and Kollias, G (2001). Uncoupling the Proinflammatory from the Immunosuppressive Properties of Tumor Necrosis Factor (TNF) at the p55 TNF Receptor Level: Implications for Pathogenesis and Therapy of Autoimmune Demyelination. *Journal of Experimental Medicine*. **193:**427-434.

Kel, A, Kel-Margoulis, O, Babenko, V and Wingender, E (1999). Recognition of NFATp/AP-1 Composite Elements within Genes Induced upon the Activation of Immune Cells. *Journal of Molecular Biology*. **288:**353-376.

Kissinger, CR, Parge, HE, Knighton, DR, Lewis, CT, Pelletier, LA, Tempczyk, A, Kalish, VJ, Tucker, KD, Showalter, RE, Moomaw, EW, Gastinel, LN, Habuka, N, Chen, X, Maldonado, F, Barker, JE, Bacquet, R and Villafranca, JE (1995). Crystal structures of human calcineurin and the human FKBP12-FK506-calcineurin complex. *Nature*. **378:**641-644.

Kleiboeker, SB, Burrage, TG, Scoles, GA, Fish, D and Rock, DL (1998). African Swine Fever Virus Infection in the Argasid Host, *Ornithodoros porcinus porcinus*. *Journal of Virology*. **72:**1711-1724.

Kleiboeker, SB and Scoles, GA (2001). Pathogenesis of African swine fever virus in *Ornithodoros* ticks. *Animal Health Research Reviews*. **2(2):**121-128.

Klement, JF, Rice, NR, Car, BD, Abbondanzo, SJ, Powers, GD, Bhatt, G, Chen, C-H, Rosen, CA and Stewart, CL (1996). IκBα deficiency results in a sustained NF-κB response and severe widespread dermatitis in mice. *Molecular and Cellular Biology*. **16:**2341-2349.

Klingenhoff, A, Frech, K, Quandt, K and Werner, T (1999). Functional promoter modules can be detected by formal models of overall nucleotide sequence similarity. *Bioinformatics*. **15:**180-186.

Knight, JC and Kwiatkowski, D (1999). Inherited Variability of Tumor Necrosis Factor Production and Susceptibility to Infectious Disease. *Proceedings of the Association of American Physicians*. **111:**290-298.

Knight, JC, Udalova, I, Hill, AVS, Greenwood, BM, Peshu, N, Marsh, K and Kwiatkowski, D (1999). A polymorphism that affects OCT-1 binding to the *TNF* promoter region is associated with severe malaria. *Nature Genetics*. **22:**145-150.

Kollias, G and Kontoyiannis, D (2002). Role of TNF/TNFR in autoimmunity: specific TNF receptor blockade may be advantageous to anti-TNF treatments. *Cytokine & Growth Factor Reviews*. **13:**315-321.

Köntgen, F, Grumont, RJ, Strasser, A, Metcalf, D, Li, R, Tarlinton, D and Gerondakis, S (1995). Mice lacking the c-*rel* proto-oncogene exhibit defects in lymphocyte proliferation, humoral immunity, and interleukin-2 expression. *Genes & Development*. **9:**1965-1977.

Kopp, EB and Ghosh, S (1995). NF-kB and Rel proteins in innate immunity. *Advances in Immunology*. **58:**1-27.

Kostavasili, J, Sahu, A, Friedman, HM, Eisenberg, RJ, Cohen, GH and Lambris, JD (1997). Mechanism of Complement Inactivation by Glycoprotein C of Herpes Simplex Virus. *Journal of Immunology*. **158:**1763-1771.

Kovalenko, YR and Sidorov, MA (1973). Reservation and Mechanisms of the Circulation of the Virus of African Hog Cholera in Nature. *Sel'skokhozyaistvennaya Biologiya*. **8:**598-606.

Krajcsi, P and Wold, WSM (1998). Inhibition of tumor necrosis factor and interferon triggered responses by DNA viruses. *Seminars in Cell & Developmental Biology*. **9(351-358)**.

Kuhnert, P, Peterhans, E and Pauli, U (1992). Chromatin structure and DNAse I hypersensitivity in the transcriptionally active and inactive porcine tumor necrosis factor gene locus. *Nucleic Acids Research*. **20:**1943-1948.

Kuhnert, P, Wüthrich, C, Peterhans, E and Pauli, U (1991). The porcine tumour necrosis factor-encoding genes: sequence and comparative analysis. *Gene*. **102:**171-178.

Kumar, S, Tamura, K, Jakobsen, IB and Nei, M (2001). MEGA2: *Molecular Evolutionary Genetics Analysis* software (version 2.1). http://www.megasoftware.net/. *Bioinformatics*. **17:**1244-1245.

Kurane, I and Takasaki, T (2001). Dengue fever and dengue haemorrhagic fever: challenges of controlling an enemy still at large. *Reviews in Medical Virology*. **11:**301-311.

Kwon, J, Lee, SJ and Benveniste, EN (1996). A 3' cis-acting element is involved in tumor necrosis factor-alpha gene expression in astrocytes. *Journal of Biological Chemistry*. **271:**22383-22390.

Lee, JY, Kim, NA, Sanford, A and Sullivan, KE (2003). Histone acetylation and chromatin conformation are regulated separately at the TNF-alpha promoter in monocytes and macrophages. *Journal of Leukocyte Biology*. **73:**862-871.

Lenardo, MJ, Chan, FK-M, Hornung, F, McFarland, H, Siegel, R, Wang, J and Zheng, L (1999). Mature T Lymphocyte Apoptosis - Immune Regulation in a Dynamic and Unpredictable Antigenic Environment. *Annual Review of Immunology*. **17:**221-253.

Leung, TH, Hoffmann, A and Baltimore, D (2004). One nucleotide in a kappaB site can determine cofactor specificity for NF-kappaB dimers. *Cell*. **118:**453-464.

Li, Q, Van Antwerp, D, Mercurio, F, Lee, K-F and Verma, IM (1999a). Severe liver degeneration in mice lacking the IκB kinase 2 gene. *Science*. **284:**321-325.

Li, Z-W, Chu, W, Hu, Y, Delhase, M, Deerinck, T, Ellisman, M, Johnson, R and Karin, M (1999b). The IKKβ subunit of IκB kinase (IKK) is essential for NF-κB activation and prevention of apoptosis. *Journal of Experimental Medicine*. **189:**1839-1845.

Liang, J, Jaffrey, SR, Guo, W, Snyder, SH and Clardy, J (1999). Structure of the PIN/LC8 dimer with a bound peptide. *Nature Structural Biology*. **6:**735-740.

Locksley, RM, Killeen, N and Lenardo, MJ (2001). The TNF and TNF Receptor Superfamilies: Integrating Mammalian Biology. *Cell*. **104:**487-501.

Lowden, S, Finlayson, HA, Macdonald, AA, Downing, AC, Goodman, SJ, Leus, K, Kaspe, L, Wahyuni, E and Archibald, AL (2002). Application of *Sus scrofa* microsatellite markers to wild suiformes. *Conservation Genetics*. **3:**347-350.

Lowden, S, Goodman, SJ, Palgrave, CJ, Tay, W and Hopwood, PA (2004). A concise molecular phylogeny of the Suidae (provisional title). *In Preparation*.

Macdonald, AA (1993). The babirusa (*Babyrousa babyrussa*). *In:* Pigs, Peccaries and Hippos. Oliver, WLR (Ed.). Gland, IUCN. pp 161-171.

MacEwan, DJ (2002). TNF ligands and receptors - a matter of life and death. *British Journal of Pharmacology*. **135:**855-875.

Macian, F, Lopez-Rodruigues, C and Rao, A (2000). Partners in transcription: NFAT and AP-1. *Oncogene*. **20:**2476-2489.

Macián, F, López-Rodruíguez, C and Rao, A (2000). Partners in transcription: NFAT and AP-1. *Oncogene*. **20:**2476-2489.

Magez, S, Truyens, C, Merimi, M, Radwanska, M, Stijlemans, B, Brouckaert, P, Brombacher, F, Pays, E and De Baetselier, P (2004). P75 tumor necrosis factor-receptor shedding occurs a protective host response during African trypanosomaiasis. *Journal of Infectious Diseases*. **189:**527-539.

Manso Ribeiro, J and Rosa Azevedo, JA (1961). Réapparition de la Peste Porcine Africaine (P. P. A.) au Portugal. *Bulletin de l'Office Internationale des Épizooties*. **55:**88-106.

Manso Ribeiro, J, Rosa Azevedo, JA, Texeira, MJO, Braco Forte, MC, Rodrigues Ribeiro, AM, Oliveira, E, Noronha, F, Grave Pereira, C and Dias Viagrio, J (1958). Peste porcine provoquée par une souche différente (Souche L) de la souche classique. *Bulletin de l'Office Internationale des Épizooties*. **50:**516-534.

Mansvelt, PR (1963). The incidence and control of African swine fever in the Republic of South Africa. *Bulletin de l'Office Internationale des Épizooties*. **60:**889-894.

Marianneau, P, Flamand, M, Deubel, V and Deprès, P (1998). Apoptotic cell death in response to dengue virus infection: the pathogenesis of dengue haemorrhagic fever revisited. *Clinical and Diagnostic Virology*. **10:**113-119.

Marta, RF, Montero, VS, Hack, CE, Sturk, A, Maiztegui, JI and Molinas, FC (1999). Proinflammatory Cytokines and Elastase-α-1-Antitrypsin in Argentine Hemorrhagic Fever. *American Journal of Tropical Medicine and Hygiene*. **60:**85-89.

Matthee, CA, Burzlaff, JD, Taylor, JF and Davis, SK (2001). Mining the Mammalian Genome for Artiodactyl Systematics. *Systematic Biology*. **50:**367-390.

Maurer, FD, Griesemer, RA and Jones, TC (1958). The Pathology of African Swine Fever - A Comparison with Hog Cholera. *American Journal of Veterinary Research*. **19:**517-539.

McDaniel, HA (1978). African Swine Fever. *American Association of Veterinary Laboratory Diagnosticians - 21st Annual Proceedings*: pp 391-398.

Mebus, CA, McVicar, JW and Dardiri, AH (1983). Comparison of the Pathology of high and Low Virulence African Swine Fever Virus Infections. *In:* CEC/FAO Expert Consultation on African Swine Fever Research, Sardinia 23-25 September 1981. EUR 8466 EN. Wilkinson, PJ (Ed.). Luxembourg, Commission of the European Communities. pp 183-194.

Mellor, PS, Kitching, RP and Wilkinson, PJ (1987). Mechanical transmission of capripox virus and African swine fever virus by *Stomoxys calcitrans*. *Research in Veterinary Science*. **43:**109-112.

Miskin, JE, Abrams, CC and Dixon, LK (2000). African swine fever virus protein A238L interacts with the cellular phosphatase calcineurin via a binding domain similar to that of NFAT. *Journal of Virology*. **74:**9412-9420.

Miskin, JE, Abrams, CC, Goatley, LC and Dixon, LK (1998). A viral mechanism for inhibition of the cellular phosphatase calcineurin. *Science*. **281:**562-565.

Monath, TP (1986). Socioeconomic Impact of Arbovirus Diseases - A World View. *Arbovirus Research in Australia, Proceedings of the Fourth Symposium*:3-16.

Montgomery, RE (1921). On a form of swine fever occurring in British East Africa (Kenya colony). *Journal of Comparative Pathology*. **34:**159-191, 243-262.

Morgan, AJ and Jacob, R (1994). Ionomycin enhances Ca^{2+} influx by stimulating store-regulated cation entry and not by direct action at the plasma membrane. *Biochemical Journal*. **300:**665-672.

Mori, M, Rothman, AL, Kurane, I, Montoya, JM, Nolte, KB, Norman, JE, Waite, DC, Koster, FT and Ennis, FA (1999). High Levels of Cytokine-Producing Cells in the Lung Tissues of Patients with Fatal Hantavirus Pulmonary Syndrome. *Journal of Infectious Diseases*. **179:**295-302.

Moulton, J and Coggins, L (1968). Comparison of Lesions in Acute and chronic African Swine Fever. *Cornell Veterinarian*. **58:**364-388.

Moura Nunes, JF, Vigário, JD and Terrinha, AM (1975). Ultrastructural Study of African Swine Fever Virus Replication in Cultures of Swine Bone Marrow Cells. *Archives of Virology*. **49:**59-66.

Neben, K, Mytilineos, J, Moehler, TM, Preiss, A, Kraemer, A, Ho, AD, Opelz, G and Goldschmidt, H (2002). Polymorphisms of the tumor necrosis factor-α gene promoter predict for outcome after thalidomide therapy in relapsed and refractory multiple myeloma. *Blood*. **100:**2263-2265.

Neilan, JG, Lu, Z, Afonso, CL, Kutish, GF, Sussman, MD and Rock, DL (1993). An African Swine Fever Virus Gene with Similarity to the Proto-Oncogene *bcl-2* and the Epstein-Barr Virus Gene *BHRF1*. *Journal of Virology*. **67:**4391-4394.

Neilan, JG, Lu, Z, Kutish, GF, Zsak, L, Lewis, TL and Rock, DL (1997). A Conserved African Swine Fever Virus IκB Homolog, 5EL, Is Nonessential for Growth *in Vitro* and Virulence in Domestic Swine. *Virology*. **235:**377-385.

Neser, JA, Phillips, T, Thomson, GR, Gainaru, MD and Coetzee, T (1986). African Swine Fever. I. Morphological Changes and Virus Replication in Blood Platelets of Pigs Infected with Virulent Haemadsorbing and Non-Haemadsorbing Isolates. *Onderstepoort Journal of Veterinary Research*. **53:**133-141.

Nogal, ML, González de Buitrago, G, Rodríguez, C, Cubelos, B, Carrascosa, AL, Salas, ML and Revilla, Y (2001). African Swine Fever IAP Homologue Inhibits Caspase Activation and Promotes Cell Survival in Mammalian Cells. *Journal of Virology*. **75:**2535-2543.

O' Leary, MA and Geisler, JH (1999). The position of Cetecae within Mammalia: Phylogenetic analysis of morphological data from extinct and extant taxa. *Systematic Biology*. **48:**455-490.

Okomo-Assoumou, MC, Daulouede, S, Lemesre, JL, N' Zila-Mouanda, A and Vincendeau, P (1995). Correlation of high serum levels of tumor necrosis factor-alpha with disease severity in human African trypanosomiasis. *American Journal of Tropical Medicine and Hygiene*. **53:**539-543.

Old, LJ (1985). Tumor necrosis factor (TNF). *Science*. **230:**630-632.

Oura, CAL, Powell, PP, Anderson, E and Parkhouse, RME (1998a). The pathogenesis of African swine fever in the resistant bush pig. *Journal of General Virology*. **79:**1439-1443.

Oura, CAL, Powell, PP and Parkhouse, RME (1998b). African swine fever: a disease characterised by apoptosis. *Journal of General Virology*. **79:**1427-1438.

Palgrave, CJ (2001). Utilising *Cytochrome b* Sequence for Basic Phylogenetic Studies. *BSc (Vet Sc) Hons* Dissertation. Faculty of Veterinary Medicine, University of Edinburgh.

Pan, IC (1987). Spontaneously Susceptible Cells and Cell Culture Methodologies for African Swine Fever Virus. *In:* African Swine Fever. Becker, Y (Ed.). Boston, Martinus Nijhoff Publishing. pp 81-126.

Pan, IC and Hess, WR (1984). Virulence in African swine fever: its measurement and implications. *American Journal of Veterinary Research*. **45(2):**361-366.

Perkins, ND (2004). NF-κB: tumor promoter or suppressor? *Trends in Cell Biology*. **14:**64-69.

Perkins, ND, Felzien, LK, Betts, JC, Leung, K, Beach, DH and Nabel, GJ (1997). Regulation of NF-kappaB by cyclin-dependent kinases associated with the p300 coactivator. *Science*. **275:**523-527.

Pfeffer, K (2003). Biological functions of tumor necrosis factor cytokines and their receptors. *Cytokine & Growth Factor Reviews*. **14:**185-191.

Plowright, W (1977). Vector Transmission of African Swine Fever Virus. *In:* Agricultural Research Seminar on Hog Cholera/Classical Swine Fever and African Swine Fever EUR 5904 EN. Leiss, BC (Ed.). Luxembourg, Commission of the European Communities. pp 575-587.

Plowright, W, Parker, J and Pierce, MA (1969a). African Swine Fever Virus in Ticks (*Ornithodoros moubata*, Murray) collected from Animal Burrows in Tanzania. *Nature (London)*. **221:**1071-1073.

Plowright, W, Parker, J and Pierce, MA (1969b). The epizootiology of African swine fever in Africa. *Veterinary Record*. **85:**668-674.

Plowright, W, Perry, CT and Greig, A (1974). Sexual Transmission of African Swine Fever Virus in the Tick, *Ornithodoros moubata porcinus*, Walton. *Research in Veterinary Science*. **17:**106-113.

Plowright, W, Perry, CT and Pierce, MA (1970). Transovarial Infection with African Swine Fever Virus in the Argasid Tick, *Ornithodoros moubata porcinus*, Walton. *Research in Veterinary Science*. **11:**582-584.

Plowright, W, Thomson, GR and Neser, JA (1994). African Swine Fever Virus. *In:* Infectious Diseases of Livestock - with Special Reference to Southern Africa (Volume One). Coetzer, JAW, Thomson, GR and Tustin, RC (Eds.). Cape Town, Oxford University Press Southern Africa. pp 567-599.

Polo Jover, F and Sanchez Botija, C (1961). La Peste Porcine Africana en España. *Bulletin de l'Office Internationale des Épizooties*. **55:**107-147.

Powell, PP, Dixon, LK and Parkhouse, RME (1996). An IκB Homolog Encoded by African Swine Fever Virus Provides a Novel Mechanism for Downregulation of Proinflammatory Cytokine Responses in Host Macrophages. *Journal of Virology*. **70(12):**8527-8533.

Pringle, CR (1999). Virus Taxonomy 1999. The Universal System of Virus Taxonomy, updated to include the new proposals ratified by the International Committee on Taxonomy of Viruses during 1998. *Archives of Virology*. **144 (2):**421-429.

Puthalakath, H, Huang, DCS, O'Reilly, LA, King, SM and Strasser, A (1999). The Proapoptotic Activity of the Bcl-2 Family Member Bim Is Regulated by Interaction with the Dynein Motor Complex. *Molecular Cell*. **3:**287-296.

Quandt, K, Frech, K, Karas, H, Wingender, E and Werner, T (1995). MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data. *Nucleic Acids Research*. **23:**4878-4884.

Ramiro-Ibáñez, F, Ortega, A, Brun, A, Escribano, JM and Alonso, C (1996). Apoptosis: a mechanism of cell killing and lymphoid organ impairment during acute African swine fever virus infection. *Journal of General Virology*. **77:**2209-2219.

Randi, E, d'Huart, J-P, Lucchini, V and Aman, R (2002). Evidence of two genetically deeply divergent species of warthog, *Phacochoerus africanus* and *P. aethiopicus* (Artiodactyla: Suiformes) in East Africa. *Mammalian Biology*. **67:**91-96.

Randi, E, Lucchini, V and Hoong Diong, C (1996). Evolutionary Genetics of the Suiformes as Reconstructed Using mtDNA Sequencing. *Journal of Mammalian Evolution*. **3(2):**163-194.

Rao, A, Luo, C and Hogan, PG (1997). Transcription Factors of the NFAT Family: Regulation and Function. *Annual Review of Immunology*. **15:**707-747.

Rennie, L, Wilkinson, PJ and Mellor, PS (2001). Transovarial transmission of African swine fever virus in the argasid tick *Ornithodoros moubata*. *Medical and Veterinary Entomology*. **15:**140-146.

Revilla, Y, Callejo, M, Rodríguez, JM, Culebras, E, Nogal, ML, Salas, ML, Viñuela, E and Fresno, M (1998). Inhibition of nuclear factor kB activation by a virus-encoded IkB-like protein. *Journal of Biological Chemistry*. **273(9):**5405-5411.

Rockman, MV, Hahn, MW, Soranzo, N, Goldstein, DB and Wray, GA (2003). Positive Selection on a Human-Specific Transcription Factor Binding Site Regulating *IL4* Expression. *Current Biology*. **13:**2118-2123.

Rodríguez, CI, Nogal, ML, Carrascosa, AL, Salas, ML, Fresno, M and Revilla, Y (2002). African Swine Fever Virus IAP-Like Protein Induces the Activation of Nuclear Factor Kappa B. *Journal of Virology*. **76:**3936-3942.

Rodriguez, F, Ley, V, Gómez-Puertas, P, García, R, Rodriguez, JF and Escribano, JM (1996). The structural protein p54 is essential for African swine fever virus viability. *Virus Research*. **40:**161-167.

Roeder, PL, Masiga, WN, Rossiter, PB, Paskin, RD and Obi, TU (1999). Dealing with animal disease emergencies in Africa: prevention and preparedness. *Revue Scientifique et Technique de l'Office International des Épizooties*. **18(1):**59-65.

Roff, M, Thompson, J, Rodriguez, MS, Jacque, J-M, Baleux, F, Arenzana-Seisdedos, F and Hay, RT (1996). Role of IκBα Ubiquitination in Signal-induced Activation of NF-κB *in Vivo*. *Journal of Biological Chemistry*. **271:**7844-7850.

Rothman, AL and Ennis, FA (1999). Immunopathogenesis of Dengue Hemorrhagic Fever. *Virology*. **257:**1-6.

Rouiller, I, Brookes, SM, Hyatt, AD, Windsor, M and Wileman, T (1998). African swine fever virus is wrapped by the endoplasmic reticulum. *Journal of Virology*. **72:**2327-2387.

Rousseau, P, Le Discorde, M, Mouillot, G, Marcou, C, Carosella, ED and Moreau, P (2003). The 14 bp deletion-insertion polymorphism in the 3' UT region of the HLA-G gene influences HLA-G mRNA stability. *Human Immunology*. **64:**1005-1010.

Rusnak, F and Mertz, P (2000). Calcineurin: Form and Function. *Physiological Reviews*. **80(4):**1483-1521.

Sakurai, H, Chiba, H, Miyoshi, H, Sugita, T and Toriumi, W (1999). IκB Kinases Phosphorylate NF-κB p65 Subunit on Serine 536 in the Transactivation Domain. *Journal of Biological Chemistry*. **274:**30353-30356.

Salguero, FJ, Ruiz-Villamor, E, Bautista, MJ, Sánchez-Cordón, PJ, Carrasco, L and Gómez-Villamandos, JC (2002). Changes in macrophages in spleen and lymph nodes during acute African swine fever: expression of cytokines. *Veterinary Immunology and Immunopathology*. **90:**11-22.

Sánchez Botija, AC (1963). Reservoirs of ASFV: A study of the ASFV in arthropods by means of haemadsorption. *Bulletin de l'Office Internationale des Épizooties*. **60:**895-899.

Sanchez Botija, C and Badiola, C (1966). Prescencia del virus de la peste Porcina Africana en el *Haematopinus suis*. *Bulletin de l'Office Internationale des Épizooties*. **66:**699-705.

Sánchez-Torrez, C, Gómez-Puertas, P, Gómez del Moral, M, Alonso, F, Escribano, JM, Ezquerra, A and Domínguez, J (2003). Expression of porcine CD163 on monocytes/macrophages correlates with permissiveness to African swine fever infection. *Archives of Virology*. **148:**2307-2323.

Sanchez-Vizcaino, JM (1999). African Swine Fever. *In:* Diseases of Swine, Eighth Edition. Straw, BE, D'Allaire, S, Mengeling, WL and Taylor, DJ (Eds.). Oxford, Blackwell Science. pp 93-102.

Sato, M, Mikami, O, Kobayashi, M and Nakajima, Y (2000). Apoptosis in the lymphatic organs of piglets inoculated with classical swine fever virus. *Veterinary Microbiology*. **75:**1-9.

Schaefer, E (2004). Ionomycin and Phorbol Ester Signalling Pathway. http://www.biocarta.com/pathfiles/ionPathway.asp#contributors.

Scherf, M, Klingenhoff, A and Werner, T (2000). Highly Specific Localization of Promoter Regions in Large Genomic Sequences by PromoterInspector: A Novel Context Analysis Approach. *Journal of Molecular Biology*. **297:**599-606.

Schloer, GM (1985). Polypeptides and structure of African swine fever virus. *Virus Research*. **3:**295-310.

Schmitz, ML, Bacher, S and Kracht, M (2001). IκB-independent control of NF-κB activity by modulatory phosphorylations. *Trends in Biochemical Sciences*. **26(3):**186-190.

Schmitz, ML, dos Santos Silva, MA and Baeuerle, PA (1995). Transactivation Domain 2 (TA_2) of p65 NF-κB. *Journal of Biological Chemistry*. **270:**15576-15584.

Schreiber, SL and Crabtree, GR (1992). The mechanism of action of cyclosporin A and FK506. *Immunology Today*. **13(4):**136-142.

Sha, WC, Liou, H-C, Tuomanen, EI and Baltimore, D (1995). Targeted disruption of the p50 subunit of NF-κB leads to multifocal defects in immune responses. *Cell*. **80:**321-330.

Sierra, AM, Bernabe, A, Mozos, E, Mendez, A and Jover, A (1987). Ultrastructure of the Liver in Pigs with Experimental African Swine Fever. *Veterinary Pathology*. **24:**460-462.

Sierra, MA, Quezada, M, Fernandez, A, Carrasco, L, Gomez-Villamandos, JC, Martin de las Mulas, J and Sanchez-Vizcaino, JM (1989). Experimental African Swine Fever: Evidence of the Virus in Interstitial Tissues of the Kidney. *Veterinary Pathology*. **26:**173-176.

Simeonidis, S, Stauber, D, Chen, G, Hendrickson, WA and Thanos, D (1999). Mechanisms by which IκB proteins control NF-κB activity. *Proceedings of the National Academy of Sciences of the United States*. **96:**49-54.

Simoons, FJ (1953). Notes on the Bushpig (*Potamochoerus*). *Uganda Journal*. **17:**80-81.

Skinner, JD and Smithers, RHN (1990). The Mammals of the Southern African Subregion. Pretoria, University of Pretoria.

Solinas, S, Pauli, U, Kuhnert, P, Peterhans, E and Fries, R (1992). Assignment of the porcine tumor necrosis factor alpha and beta genes to the chromosome region 7p11-q11 by *in situ* hybridization. *Animal Genetics*. **23:**267-271.

Staden, R, Beal, KF and Bonfield, JK (1999). The Staden Package, 1998. Computer Methods in Molecular Biology (http://www.mrc-lmb.cam.ac.uk/pubseq/2002.0.html). *In:* Bioinformatics Methods and Protocols. Misener, S and Krawetz, SA (Eds.). Totowa, The Humana Press Inc. pp 115-130.

Staden, R, Bonfield, JK, Beal, KF, Cheng, Y and Jordan, M (2002). The Staden Package. *http://www.mrc-lmb.cam.ac.uk/pubseq/2002.0.html*.

Stevenson, PG, Efstathiou, S, Doherty, PC and Lehner, PJ (2000). Inhibition of MHC class I-restricted antigen presentation by γ2-herpesviruses. *Proceedings of the National Academy of Sciences of the United States*. **97:**8455-8460.

Steyn, DG (1928). Preliminary Report on a South African Virus Disease amongst Pigs. *In:* 13th and 14th Reports of the Director of Veterinary Education and Research. Union of South Africa. pp 415-428.

Steyn, DG (1932). East African Virus Disease in Pigs. *In:* 18th Report of the Director of Veterinary Services and Animal Industry. Union of South Africa. pp 99-109.

Stone, SS and Heuschele, WP (1965). The Role of the Hippopotamus in the Epizootiology of African Swine Fever. *Bulletin of Epizootic Diseases in Africa*. **13:**23-28.

Sulandari, S, Muladno, Harumi, T, Yanai, S, Wada, Y and Yasue, H (1997). Localization of swine PRE-1 homologues in 13 loci of Phacochoerus aethiopicus and Tayassu tajacu genomes, and their sequence divergence. *Animal Genetics*. **28:**210-215.

Tait, SWG, Reid, EB, Greaves, DR, Wileman, TE and Powell, PP (2000). Mechanism of Inactivation of NF-κB by a Viral Homologue of IκBα. *The Journal of Biological Chemistry*. **275:**34656-34664.

Takamatsu, H, Denyer, MS, Oura, C, Childerstone, A, Anderson, JK, Pullen, L and Parkhouse, RME (1999). African swine fever virus: a B cell-mitogenic virus in vivo and in vitro. *Journal of General Virology*. **80:**1453-1461.

Takeda, K, Takeuchi, O, Tsujimura, T, Itami, S, Adachi, O, Kawai, T, Sanjo, H, Yoshikawa, K, Terada, N and Akira, S (1999). Limb and Skin Abnormalities in Mice Lacking IKKα. *Science*. **284:**313-316.

Tanaka, M, Fuentes, ME, Yamaguchi, K, Durnin, MH, Dalrymple, SA, Hardy, KL and Goedel, DV (1999). Embryonic lethality, liver degeneration, and impaired NF-κB activation in IKK-β-deficient mice. *Immunity*. **10:**421-429.

Thenius, E (1970). Zur Evolution und Verbreitungsgeschichte der Suidae (Artiodactyla, Mammalia). *Zeitschrift für Säugertierkunde*. **35:**321-342.

Thiagalingam, S, Cheng, K-H, Lee, HJ, Mineva, N, Thiagalingam, A and Ponte, JF (2003). Histone Deacetylases: Unique Players in Shaping the Epigenetic Code. *Annals of the New York Academy of Sciences*. **983:**84-100.

Thomson, GR (1985). The epidemiology of African swine fever: the role of free-living hosts in Africa. *Onderstepoort Journal of Veterinary Research*. **52:**201-209.

Thomson, GR, Gainaru, MD and Van Dellen, AF (1980). Experimental infection of warthog (*Phacochoerus aethiopicus*) with African swine fever virus. *Onderstepoort Journal of Veterinary Research*. **47:**19-22.

Tizard, I (1996). Veterinary Immunology. Philadelphia, W.B. Saunders.

Tochio, H, Ohki, S, Zhang, Q, Li, M and Zhang, M (1998). Solution structure of a protein inhibitor of neuronal nitric oxide synthase. *Nature Structural Biology*. **5:**965-969.

Tulman, ER and Rock, DL (2001). Novel virulence and host range genes of African swine fever virus. *Current Opinion in Microbiology*. **4:**456-461.

Turner, BM (2002). Cellular Memory and the Histone Code. *Cell*. **111:**285-291.

Vallée, I, Tait, SWG and Powell, PP (2001). African Swine Fever Virus Infection of Porcine Endothelial Cells Leads to Inhibition of Inflammatory Responses, Activation of the Thrombotic State, and Apoptosis. *Journal of Virology*. **75:**10372-10382.

Vercammen, P and Mason, DR (1993). The warthogs (*Phacochoerus africanus* and *P. aethiopicus*). *In:* Pigs, Peccaries, and Hippos. Oliver, WLR (Ed.). Gland, IUCN. pp 75-84.

Vercammen, P, Seydack, AHW and Oliver, WLR (1993). The Bush Pigs (*Potamochoerus larvatus* and *P. porcus*). *In:* Pigs, Peccaries, and Hippos. Oliver, WLR (Ed.). Gland, IUCN. pp 93-101.

Vermeulen, L, De Wilde, G, Notebaert, S, Berghe, WV and Haegeman, G (2002). Regulation of the transcriptional activity of the nuclear factor-κB p65 subunit. *Biochemical Pharmacology*. **64:**963-970.

Villinger, F, Rollin, PE, Brar, SS, Chikkala, NF, Winter, J, Sundstrom, JB, Zaki, SR, Swanepoel, R, Ansari, AA and Peters, CJ (1999). Markedly Elevated Levels of Interferon (IFN)-γ, IFN-α, Interleukin (IL)-2, IL-10, and Tumor Necrosis Factor-α Associated with Fatal Ebola Virus Infection. *Journal of Infectious Diseases*. **179(Suppl. 1):**S188-S191.

Viñuela, E (1985). African Swine Fever Virus. *Current Topics in Microbiology and Immunology*. **116:**151-170.

Viñuela, E (1987). Molecular Biology of African Swine Fever Virus. *In:* African Swine Fever. Becker, Y (Ed.). Boston, Martinus Nijhoff Publishing. pp 31-49.

Vyas, P, Vickers, MA, Picketts, DJ and Higgs, DR (1995). Conservation of position and sequence of a novel, widely expressed gene containing the major human alpha-globin regulatory element. *Genomics*. **29:**679-689.

Wajant, H, Pfizenmaier, K and Scheurich, P (2003). Tumor necrosis factor signaling. *Cell Death and Differentiation*. **10:**45-65.

Wang, D and Baldwin Jr., AS (1998). Activation of Nuclear Factor-κB-dependent Transcription by Tumor Necrosis Factor-α Is Mediated through Phosphorylation of RelA/p65 on Serine 529. *Journal of Biological Chemistry*. **273:**29411-29416.

Wang, D, Westerheide, SD, Hanson, JL and Baldwin Jr., AS (2000). Tumor Necrosis Factor α-induced Phosphorylation of RelA/p65 on Ser^{529} Is Controlled by Casein Kinase II. *Journal of Biological Chemistry*. **275:**32592-32597.

Wardley, RC, Andrade, CdM, Black, DN, de Castro Portugal, FL, Enjuanes, L, Hess, WR, Mebus, C, Ordans, A, Rutili, D, Sanchez-Vizcaino, JM, Vigario, JD and Wilkinson, PJ (1983). African Swine Fever Virus - Brief Review. *Archives of Virology*. **76:**73-90.

Wardley, RC and Wilkinson, PJ (1977). The Association of African Swine Fever Virus With Blood Components of Infected Pigs. *Archives of Virology*. **55:**327-334.

Ware, CF (2003). The TNF Superfamily. *Cytokine & Growth Factor Reviews*. **14:**181-184.

Weih, F, Carrasco, D, Durham, SK, Barton, DS, Rizzo, CA, Ryseck, R-P, Lira, SA and Bravo, R (1995). Multiorgan inflammation and haematopoietic abnormalities in mice with a targeted disruption of RelB. *Cell*. **80:**331-340.

Weih, F, Durham, SK, Barton, DS, Sha, WC, Baltimore, D and Bravo, R (1997). p50-NF-κB complexes partially compensate for the absence of RelB: Severely increased pathology in *p50*$^{-/-}$ *relB*$^{-/-}$ double-knockout mice. *Journal of Experimental Medicine*. **185:**1359-1370.

Whitelaw, CB (2000). Nucleosome organisation of the beta-lactoglobin gene. Transcription complex formation. *Advances in Experimental Medicine and Biology*. **480:**147-153.

Wilkinson, PJ (1989). African swine fever. *In:* Virus Infections in Porcines. Pensaert, MB (Ed.). Amsterdam, Elsevier. pp 17-35.

Wilkinson, PJ, Lawman, MJP and Johnston, RS (1980). African swine fever in Malta, 1978. *Veterinary Record*. **106:**94-97.

Wilkinson, PJ and Wardley, RC (1978). The Replication of African Swine Fever Virus in Pig Endothelial Cells. *British Veterinary Journal*. **134:**280-282.

Wilson, AG, di Giovine, FS and Duff, GW (1995). Genetics of Tumour Necrosis Factor-α in Autoimmune, Infectious, and Neoplastic Diseases. *Journal of Inflammation*. **45:**1-12.

Yenidunya, A, Davey, C, Clark, D, Felsenfeld, G and Allan, J (1994). Nucleosome positioning on chicken and human globin gene promoters in vitro. Novel mapping techniques. *Journal of Molecular Biology*. **237:**401-414.

Zandi, E and Karin, M (1999). Minireview. Bridging the Gap: Composition, Regulation, and Physiological Function of the IκB Kinase Complex. *Molecular and Cellular Biology*. **19:**4547-4551.

Zhong, H, May, MJ, Jimi, E and Ghosh, S (2002). The Phosphorylation Status of Nuclear NF-κB Determines Its Association with CBP/p300 or HDAC-1. *Molecular Cell*. **9:**625-636.

Zhong, H, Voll, RE and Ghosh, S (1998). Phosphorylation of NF-κB p65 by PKA Stimulates Transcriptional Activity by Promoting a Novel Bivalent Interaction with the Coactivator CBP/p300. *Molecular Cell*. **1:**661-671.

Zhou, A, Scoggin, S, Gaynor, RB and Williams, NS (2003). Identification of NF-κB-regulated genes induced by TNFα utilizing expression profiling and RNA interference. *Oncogene*. **22:**2054-2064.

Zhuang, Y, Butler, B, Hawkins, E, Paguio, A, Orr, L, Wood, MG and Wood, KV (2001). A New Age of Enlightenment - New synthetic Renilla Gene and Assay System Increase Expression, Reliability and Sensitivity. *Promega Notes*. **79:**6-11.

8 Appendices

1. MatInspector-predicted transcription factor binding domains in the domestic pig TNFα promoter.

2. MatInspector-predicted transcription factor binding domains in the warthog TNFα promoter.

3. Additional suid TNFα promoter sequences determined for phylogenetic analysis.

Appendix 1

The 170 predicted transcription factor binding sites within the domestic pig TNFα promoter with a matrix similarity greater than 0.75. MatInspector provides a code to describe the family of each transcription factor, a brief description of the individual transcription factor, the location of the binding domain, which DNA strand (+ or -) the domain is encoded by, the matrix similarity and the nucleotide sequence of the domain. Uppercase letters denote the four most highly conserved bases (core) within the domain.

Family/matrix	Description	Position	Str.	Matrix sim.	Sequence
V$ETSF/ETS1.01	c-Ets-1 binding site	1 - 17	(+)	0.942	ggctcAGGAaggggctg
V$PBXC/PBX1_MEIS1.03	Binding site for a Pbx1/Meis1 heterodimer	12 - 28	(+)	0.805	gggctgctTGACtggag
V$MINI/MUSCLE_INI.01	Muscle Initiator Sequence	14 - 32	(-)	0.880	gagcctCCAGtcaagcagc
V$PAX2/PAX2.01	Zebrafish PAX2 paired domain protein	41 - 63	(+)	0.791	acggctgaccctcgatgAAACcc
V$PCAT/CAAT.01	Cellular and viral CCAAT box	58 - 68	(+)	0.916	aaacCCAAtaa
V$ECAT/NFY.02	Nuclear factor Y (Y-box binding factor)	58 - 72	(+)	0.910	aaacCCAAtaaagct
V$STAT/STAT5.01	STAT5: signal transducer and activator of transcription 5	71 - 89	(+)	0.938	ctcttTTCTctgaaatgct
V$STAT/STAT5.01	STAT5: signal transducer and activator of transcription 5	71 - 89	(-)	0.964	agcatTTCAgagaaaagag
V$EVI1/EVI1.04	Ecotropic viral integration site 1 encoded factor	87 - 103	(-)	0.803	aGATAcgagcagacagc
V$GATA/GATA1.01	GATA-binding factor 1	94 - 106	(-)	0.961	gacaGATAcgagc
V$SRFF/SRF.01	Serum response factor	92 - 110	(+)	0.660	ctgctcgTATCtgtcactc
V$MEIS/MEIS1.01	Binding site for monomeric Meis1 homeodomain protein	100 - 108	(-)	0.971	gtGACAgat
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	109 - 121	(+)	0.964	tcggGAGGggaga
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	113 - 119	(+)	0.985	gaGGGGa
V$SORY/HMGIY.01	HMGI(Y) high-mobility-group protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex	114 - 130	(-)	0.922	ctggagAATTctcccct
V$SORY/HMGIY.01	HMGI(Y) high-mobility-group protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex	115 - 131	(+)	0.924	ggggagAATTctccaga
V$AP4R/TAL1BETAITF2.01	Tal-1beta/ITF-2 heterodimer	123 - 139	(+)	0.854	ttctcCAGAtgtctcta
V$RP58/RP58.01	Zinc finger protein RP58 (ZNF238), associated preferentially with heterochromatin	125 - 137	(-)	0.941	gagaCATCtggag
V$HAND/HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12	127 - 141	(-)	0.776	cttagaGACAtctgg
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	165 - 171	(+)	0.985	gaGGGGa
V$AP4R/AP4.03	Activator protein 4	169 - 185	(-)	0.863	acccaCAGCtcccgtcc
V$RREB/RREB1.01	Ras-responsive element binding protein 1	171 - 185	(-)	0.791	aCCCAcagctcccgt
V$AREB/AREB6.02	AREB6 (Atp1a1 regulatory element binding factor 6)	205 - 217	(-)	0.978	ctctCACCtgagc
V$HEN1/HEN1.01	HEN1	201 - 221	(+)	0.837	aggggctCAGGtgagagatgg
V$HIFF/DEC1.01	Basic helix-loop-helix protein known as Dec1, Stra13 or Sharp2	204 - 218	(-)	0.860	tctctCACCtgagcc
V$EBOR/DELTAEF1.01	deltaEF1	205 - 219	(+)	1.000	gctcAGGTgagagat
V$RXRF/VDR_RXR.01	VDR/RXR Vitamin D receptor RXR heterodimer site	209 - 225	(+)	0.868	aggtgagaGATGgcggc

V$MINI/MUSCLE_INI.02	Muscle Initiator Sequence	223 - 241	(-)	0.867	ctgcccTCACcctgaggcc
V$AP4R/TH1E47.01	Thing1/E47 heterodimer, TH1 bHLH member specific expression in a variety of embryonic tissues	235 - 251	(+)	0.948	agggcagCCAGaccccct
V$SMAD/SMAD3.01	Smad3 transcription factor involved in TGF-beta signaling	239 - 247	(-)	0.994	GTCTggctg
V$AP4R/TAL1BETAE47.01	Tal-1beta/E47 heterodimer	257 - 273	(+)	0.953	agaagCAGAtggtcctc
V$NEUR/NEUROG.01	Neurogenin 1 and 3 (ngn1/3) binding sites	259 - 271	(-)	0.973	ggaCCATctgctt
V$HAND/HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12	261 - 275	(-)	0.795	cagaggACCAtctgc
V$EVI1/EVI1.02	Ecotropic viral integration site 1 encoded factor	269 - 285	(+)	0.903	tcctctgagAAGAcaaa
V$EVI1/EVI1.01	Ecotropic viral integration site 1 encoded factor	274 - 290	(+)	0.739	tgagAAGAcaaaggaag
V$ETSF/NRF2.01	Nuclear respiratory factor 2	280 - 296	(+)	0.877	gacaaaGGAAgagatgc
V$P53F/P53.01	Tumor suppressor p53	288 - 308	(-)	0.662	gacCTTGgccctgcatctctt
V$SF1F/FTF.01	Alpha (1)-fetoprotein transcription factor (FTF), liver receptor homologue-1 (LHR-1)	298 - 310	(+)	0.952	gggcCAAGgtctt
V$RORA/NBRE.01	Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)	298 - 314	(+)	0.895	gggccAAGGtcttgaga
V$NRSF/NRSF.01	Neuron-restrictive silencer factor	309 - 329	(+)	0.693	ttgAGAAccgaggtcgggggt
V$DEAF/NUDR.01	NUDR (nuclear DEAF-1 related transcriptional regulator protein)	319 - 337	(+)	0.732	aggTCGGgggtcgcctggc
V$SRFF/SRF.01	Serum response factor	331 - 349	(-)	0.748	gtggccaTATCtgccaggc
V$GATA/LMO2COM.02	complex of Lmo2 bound to Tal-1, E2A proteins, and GATA-1, half-site 2	335 - 347	(+)	0.989	ggcaGATAtggcc
V$SRFF/SRF.03	Serum responsive factor	332 - 350	(+)	0.797	cctggcagatATGGccaca
V$AREB/AREB6.01	AREB6 (Atp1a1 regulatory element binding factor 6)	346 - 358	(-)	0.943	cctctACCTgtgg
V$MEIS/MEIS1.01	Binding site for monomeric Meis1 homeodomain protein	374 - 382	(+)	0.960	gtGACAgga
V$STAT/STAT6.01	STAT6: signal transducer and activator of transcription 6	381 - 399	(+)	0.845	gaggcTTCCcgggagaagg
V$STAT/STAT1.01	Signal transducer and activator of transcription 1	381 - 399	(-)	0.803	ccttctcccgGGAAgcctc
V$NOLF/OLF1.01	Olfactory neuron-specific factor	381 - 403	(+)	0.823	gaggctTCCCgggagaagggaac
V$E2FF/E2F.01	E2F, involved in cell cycle regulation, interacts with Rb p107 protein	386 - 400	(+)	0.786	ttcccggGAGAaggg
V$PAX5/PAX9.01	Zebrafish PAX9 binding sites	397 - 425	(+)	0.802	agggaacaCACTgaggggtgttcgggatt
V$DEAF/NUDR.01	NUDR (nuclear DEAF-1 related transcriptional regulator protein)	415 - 433	(+)	0.757	tgtTCGGgattctgaggga
V$STAT/STAT6.01	STAT6: signal transducer and activator of transcription 6	420 - 438	(-)	0.848	gctccTCCCtcagaatccc
V$NRSF/NRSE.01	Neural-restrictive-silencer-element	430 - 450	(+)	0.687	gggaggagcaCGGGgacgccc
V$NFKB/NFKAPPAB50.01	NF-kappaB (p50)	439 - 453	(+)	0.891	acGGGGacgccctgg
V$WHZF/WHN.01	Winged helix protein, involved in hair keratinization and thymus epithelium differentiation	442 - 452	(+)	0.962	gggACGCcctg
V$P53F/P53.03	Tumor suppressor p53 (3' half	454 - 474	(-)	0.928	gccctggcacggcCATGtctc

	site)				
V$P53F/P53.01	Tumor suppressor p53	455 - 475	(+)	0.687	agaCATGgccgtgccagggcc
V$MOKF/MOK2.01	Ribonucleoprotein associated zinc finger protein MOK-2 (mouse)	461 - 481	(+)	0.779	ggccgtgccagggCCATgagg
V$HAND/HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12	467 - 481	(+)	0.775	gccaggGCCAtgagg
V$EGRF/WT1.01	Wilms Tumor Suppressor	479 - 493	(+)	0.915	aggagTGGGagagcc
V$P53F/P53.02	Tumor suppressor p53 (5' half site)	528 - 548	(+)	0.910	gaggagacacggtCATGgcca
V$P53F/P53.02	Tumor suppressor p53 (5' half site)	537 - 557	(-)	0.912	cggcgctcctggcCATGaccg
V$BCL6/BCL6.01	POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma	567 - 583	(-)	0.903	agaTTCCtagagactgt
V$ETSF/ETS2.01	c-Ets-2 binding site	588 - 604	(+)	0.905	gaagcAGGAattctttg
V$LEFF/LEF1.02	TCF/LEF-1, involved in the Wnt signal transduction pathway	595 - 611	(-)	0.954	gtatcctCAAAgaattc
V$ZBPF/ZF9.01	Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers	626 - 648	(-)	0.871	gaagtccCCACacctcagcctcc
V$MINI/MUSCLE_INI.01	Muscle Initiator Sequence	629 - 647	(-)	0.873	aagtccCCACacctcagcc
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	638 - 644	(+)	1.000	gtGGGGa
V$MINI/MUSCLE_INI.02	Muscle Initiator Sequence	641 - 659	(+)	0.879	gggactTCATgcagaagtc
V$ARP1/ARP1.01	apolipoprotein AI regulatory protein 1	652 - 668	(-)	0.826	tgaggcCCTGacttctg
V$TEAF/TEF1.01	TEF-1 related muscle factor	667 - 679	(+)	0.916	caCATTcccttgg
V$SF1F/FTF.01	Alpha (1)-fetoprotein transcription factor (FTF), liver receptor homologue-1 (LHR-1)	670 - 682	(-)	0.941	cttcCAAGgaat
V$STAT/STAT6.01	STAT6: signal transducer and activator of transcription 6	667 - 685	(+)	0.854	cacatTCCCttggaagccg
V$SMAD/SMAD4.01	Smad4 transcription factor involved in TGF-beta signaling	681 - 689	(-)	0.969	GTCTcggct
V$IRFF/IRF4.01	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)	683 - 701	(+)	0.960	ccgagactGAAAccagcag
V$MYOD/MYF5.01	Myf5 myogenic bHLH protein	692 - 706	(+)	0.920	aaacCAGCagcagag
V$ETSF/PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells	710 - 726	(-)	0.891	ctgacaGGAActcacca
V$MEIS/MEIS1.01	Binding site for monomeric Meis1 homeodomain protein	718 - 726	(-)	0.951	ctGACAgga
V$TALE/TGIF.01	TG-interacting factor belonging to TALE class of homeodomain factors	721 - 727	(+)	1.000	tGTCAga
V$BARB/BARBIE.01	Barbiturate-inducible element	727 - 741	(+)	0.916	agtgAAAGgagaagg
V$MOKF/MOK2.02	Ribonucleoprotein associated zinc finger protein MOK-2 (human)	734 - 754	(-)	0.984	ccaccatggcgggCCTTctcc
V$YY1F/YY1.01	Yin and Yang 1	741 - 759	(+)	0.848	gcccgCCATggtgggtttg
V$OCT1/OCT1.06	Octamer-binding factor 1	754 - 768	(+)	0.802	ggtttgtgAATTccc
V$FAST/FAST1.01	FAST-1 SMAD interacting protein	755 - 769	(+)	0.880	gtttgtgAATTccca
V$NFKB/NFKAPPAB.01	NF-kappaB	756 - 770	(-)	0.894	ctGGGAattcacaaa
V$SORY/HMGIY.01	HMGI(Y) high-mobility-group	755 - 771	(-)	0.920	gctgggAATTcacaaac

	protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex				
V$IKRS/IK1.01	Ikaros 1, potential regulator of lymphocyte differentiation	760 - 772	(-)	0.925	ggctGGGAattca
V$GABF/GAGA.01	GAGA-Box	764 - 788	(-)	0.797	gagggAGAGgaagccaggctgggaa
V$ETSF/PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells	770 - 786	(-)	0.874	gggagaGGAAgccaggc
V$GABF/GAGA.01	GAGA-Box	768 - 792	(-)	0.790	cccagAGGGagaggaagccaggctg
V$ZBPF/ZBP89.01	Zinc finger transcription factor ZBP-89	811 - 833	(+)	0.961	ctgccgtcctCCCCcagcccgtg
V$AP2F/AP2.01	Activator protein 2	819 - 831	(+)	0.890	ctCCCCcagcccg
V$SRFF/SRF.03	Serum responsive factor	824 - 842	(-)	0.793	gaggccctacACGGgctgg
V$AP4R/AP4.01	Activator protein 4	837 - 853	(-)	0.970	aagggCAGCtggaggcc
V$MYOD/MYF5.01	Myf5 myogenic bHLH protein	838 - 852	(-)	0.917	agggCAGCtggaggc
V$ETSF/GABP.01	GABP: GA binding protein	860 - 876	(-)	0.865	ctggagGGAAgaggagc
V$NRSF/NRSF.01	Neuron-restrictive silencer factor	893 - 913	(+)	0.734	ctcAGCAcccagcgcggtgtc
V$E2TF/E2.01	BPV bovine papilloma virus regulator E2	896 - 912	(-)	0.894	acaccgcgctGGGTgct
V$E2TF/E2.01	BPV bovine papilloma virus regulator E2	897 - 913	(+)	0.935	gcacccagcgCGGTgtc
V$CDEF/CDE.01	Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)	901 - 913	(+)	0.940	ccagCGCGgtgtc
V$NKXH/NKX31.01	Prostate-specific homeodomain protein NKX3.1	914 - 926	(+)	0.847	gtctAAGTtttct
V$MYT1/MYT1.02	MyT1 zinc finger transcription factor involved in primary neurogenesis	915 - 927	(+)	0.892	tctAAGTtttctc
V$E2FF/E2F.01	E2F, involved in cell cycle regulation, interacts with Rb p107 protein	918 - 932	(-)	0.763	atggagaGAAAactt
V$NKXH/NKX25.02	Homeo domain factor Nkx-2.5/Csx, tinman homolog low affinity sites	925 - 937	(-)	0.903	tctTAATggagag
V$HOXF/HOX1-3.01	Hox-1.3, vertebrate homeobox protein	925 - 941	(+)	0.841	ctctccATTAagaactc
V$PDX1/ISL1.01	Pancreatic and intestinal lim-homeodomain factor	923 - 943	(-)	0.826	ctgagttctTAATggagagaa
V$MYT1/MYT1L.01	Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1	937 - 949	(-)	0.949	agaaAGCTgagtt
V$SP1F/GC.01	GC box elements	951 - 965	(-)	0.910	gaatgGGAGgagctt
V$TEAF/TEF1.01	TEF-1 related muscle factor	959 - 971	(+)	0.869	ccCATTcctagtt
V$AREB/AREB6.01	AREB6 (Atp1a1 regulatory element binding factor 6)	975 - 987	(+)	0.949	cccctACCTgagc
V$GFI1/GFI1.01	Growth factor independence 1 zinc finger protein acts as transcriptional repressor	995 - 1009	(+)	0.963	ggaAATCagagagaa
V$IRFF/IRF7.01	Interferon regulatory factor 7 (IRF-7)	1005 - 1023	(+)	0.871	gaGAAAtagaagtcatccc
V$CREB/TAXCREB.01	Tax/CREB complex	1005 - 1025	(-)	0.825	gggggaTGACttctatttctc
V$ETSF/PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells	1028 - 1044	(+)	0.869	agaaaaGGAAtttgtcc
V$IRFF/IRF7.01	Interferon regulatory factor 7	1027 - 1045	(+)	0.907	aaGAAAaggaatttgtccc

	(IRF-7)				
V$HMTB/MTBF.01	Muscle-specific Mt binding site	1033 - 1041	(+)	0.931	aggaATTTg
V$RREB/RREB1.01	Ras-responsive element binding protein 1	1044 - 1058	(+)	0.837	cCCCAaagaaacaga
V$AREB/AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	1048 - 1060	(-)	0.997	gttctGTTTcttt
V$P53F/P53.02	Tumor suppressor p53 (5' half site)	1047 - 1067	(+)	0.924	caaagaaacagaaCTTGtccc
V$IRFF/IRF7.01	Interferon regulatory factor 7 (IRF-7)	1049 - 1067	(+)	0.888	aaGAAAcagaacttgtccc
V$P53F/P53.03	Tumor suppressor p53 (3' half site)	1056 - 1076	(-)	0.947	tttctttgggggaCAAGttct
V$CLOX/CDPCR3.01	Cut-like homeodomain protein	1064 - 1080	(+)	0.755	tcccccaaagaaATGGa
V$IRFF/IRF3.01	Interferon regulatory factor 3 (IRF-3)	1071 - 1089	(+)	0.950	aagaaatgGAAAcaatggg
V$OCT1/OCT1.04	Octamer-binding factor 1	1073 - 1087	(+)	0.813	gaAATGgaaacaatg
V$FKHD/FKHRL1.01	Fkh-domain factor FKHRL1 (FOXO)	1073 - 1089	(+)	0.876	gaaatggaAACAatggg
V$SORY/SOX5.01	Sox-5	1078 - 1094	(+)	0.986	ggaaaCAATgggaaatg
V$RBPF/RBPJK.02	Mammalian transcriptional repressor RBP-Jkappa/CBF1	1082 - 1096	(+)	0.951	acaaTGGGaaatggg
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	1102 - 1108	(+)	0.990	ggGGGGa
V$RREB/RREB1.01	Ras-responsive element binding protein 1	1101 - 1115	(-)	0.810	cCCCAggtcccccct
V$VMYB/VMYB.04	v-Myb, AMV v-myb	1145 - 1155	(+)	0.852	agtAACTggcc
V$CMYB/CMYB.02	c-Myb, important in hematopoesis, cellular equivalent to avian myoblastosis virus oncogene v-myb	1145 - 1157	(+)	0.966	agTAACtggccca
V$REBV/EBVR.01	Epstein-Barr virus transcription factor R	1163 - 1183	(-)	0.863	ctgcccggatcctgaGGTGgg
V$EKLF/EKLF.01	Erythroid krueppel like factor (EKLF)	1180 - 1192	(+)	0.940	gcagggaGGGTag
V$ETSF/ETS1.01	c-Ets-1 binding site	1186 - 1202	(+)	0.934	agggtAGGAagtatccc
V$EGRF/EGR1.01	Egr-1/Krox-24/NGFI-A immediate-early gene product	1205 - 1219	(+)	0.840	atgcctgGGTGtccc
V$BNCF/BNC.01	Basonuclin, cooperates with USF1 in rDNA PolI transcription)	1204 - 1222	(+)	0.890	gatgcctgggTGTCcccaa
V$NFKB/NFKAPPAB50.01	NF-kappaB (p50)	1208 - 1222	(-)	0.888	ttGGGGacacccagg
V$MYT1/MYT1L.01	Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1	1216 - 1228	(-)	0.986	ggaaAGTTgggga
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	1230 - 1244	(-)	0.919	gcgggGGCGgcggtt
V$EGRF/NGFIC.01	Nerve growth factor-induced protein C	1232 - 1246	(-)	0.875	taGCGGgggcggcgg
V$ZBPF/ZBP89.01	Zinc finger transcription factor ZBP-89	1228 - 1250	(+)	0.935	caaaccgccgCCCCcgctatgga
V$SRFF/SRF.03	Serum responsive factor	1236 - 1254	(+)	0.794	cgcccccgctATGGagatg
V$IRFF/ISRE.01	Interferon-stimulated response element	1246 - 1264	(+)	0.821	atggagatGAAActaagac
V$ETSF/ETS1.01	c-Ets-1 binding site	1282 - 1298	(-)	0.925	tctggAGGAagcggtag
V$RP58/RP58.01	Zinc finger protein RP58 (ZNF238), associated preferentially with heterochromatin	1292 - 1304	(-)	0.850	agctCATCtggag

V$ECAT/NFY.03	Nuclear factor Y (Y-box binding factor)	1300 - 1314	(-)	0.812	gaaaCCCAtgagctc
V$PCAT/ACAAT.01	Avian C-type LTR CCAAT box	1315 - 1325	(+)	0.834	tccaCCAAgga
V$ETSF/ETS1.01	c-Ets-1 binding site	1317 - 1333	(+)	0.942	caccaAGGAagttttcc
V$MYT1/MYT1.02	MyT1 zinc finger transcription factor involved in primary neurogenesis	1322 - 1334	(+)	0.895	aggAAGTtttccg
V$E2FF/E2F.02	E2F, involved in cell cycle regulation, interacts with Rb p107 protein	1324 - 1338	(-)	0.884	ccagcggAAAActtc
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	1354 - 1360	(-)	0.985	gcGGGGa
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	1352 - 1364	(-)	0.913	gaggGCGGggaga
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	1352 - 1366	(-)	0.942	aagagGGCGgggaga
V$TBPF/TATA.01	cellular and viral TATA box elements	1377 - 1393	(+)	0.966	gcgtaTAAAtgcagctg
V$MYOD/E47.01	MyoD/E47 and MyoD/E12 dimers	1384 - 1398	(+)	0.922	aatGCAGctgtttgc
V$RBPF/RBPJK.01	Mammalian transcriptional repressor RBP-Jkappa/CBF1	1412 - 1426	(-)	0.862	actcTGGGagcttct
V$GLIF/GLI1.01	Zinc finger transcription factor GLI1	1438 - 1452	(+)	0.895	gggaccaGCCAggag
V$HAML/AML3.01	Runt-related transcription factor 2 / CBFA1 (core-binding factor, runt domain, alpha subunit 1)	1455 - 1469	(-)	0.846	tggaGTGGcttgtct
V$BCL6/BCL6.02	POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma	1471 - 1487	(+)	0.772	gacccccTAGAaataac
V$MEF2/HMEF2.01	Myocyte enhancer factor	1470 - 1492	(+)	0.891	ggacccccctagAAATaacctctc
V$EKLF/BKLF.01	Basic krueppel-like factor (KLF3)	1495 - 1507	(-)	0.960	ggGGGTgtgtctt
V$GABF/GAGA.01	GAGA-Box	1518 - 1542	(-)	0.819	gtgtgAGAGggagagagtcgtccgg
V$GABF/GAGA.01	GAGA-Box	1520 - 1544	(-)	0.794	gcgtgTGAGagggagagagtcgtcc
V$AHRR/AHR.01	Aryl hydrocarbon / dioxin receptor	1532 - 1554	(-)	0.862	ccccggggcaGCGTgtgagaggg
V$AP2F/AP2.01	Activator protein 2	1544 - 1556	(-)	0.907	cgCCCCggggcag
V$AP2F/AP2.01	Activator protein 2	1545 - 1557	(+)	0.915	tgCCCCggggcgc
V$AP4R/AP4.02	Activator protein 4	1563 - 1579	(+)	0.970	tctccCAGCtggacctg
V$HAND/HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12	1567 - 1581	(-)	0.778	ctcaggTCCAgctgg
V$PAX6/PAX6.02	PAX6 paired domain and homeodomain are required for binding to this site	1565 - 1583	(-)	0.908	ggctcaggtCCAGctggga

Appendix 2

The 173 predicted transcription factor binding sites within the warthog TNFα promoter with a matrix similarity greater than 0.75. MatInspector provides a code to describe the family of each transcription factor, a brief description of the individual transcription factor, the location of the binding domain, which DNA strand (+ or -) the domain is encoded by, the matrix similarity and the nucleotide sequence of the domain. Uppercase letters denote the four most highly conserved bases (core) within the domain.

Family/matrix	Description	Position	Str.	Matrix sim.	Sequence
V$ETSF/ETS1.01	c-Ets-1 binding site	1 - 17	(+)	0.942	ggctcAGGAaggggctg
V$PBXC/PBX1_MEIS1.03	Binding site for a Pbx1/Meis1 heterodimer	12 - 28	(+)	0.805	gggctgctTGACtggag
V$MINI/MUSCLE_INI.01	Muscle Initiator Sequence	14 - 32	(-)	0.880	gagcctCCAGtcaagcagc
V$PAX2/PAX2.01	Zebrafish PAX2 paired domain protein	41 - 63	(+)	0.791	acggctgaccctcgatgAAACcc
V$PCAT/CAAT.01	Cellular and viral CCAAT box	58 - 68	(+)	0.916	aaacCCAAtaa
V$ECAT/NFY.02	Nuclear factor Y (Y-box binding factor)	58 - 72	(+)	0.910	aaacCCAAtaaagct
V$STAT/STAT5.01	STAT5: signal transducer and activator of transcription 5	71 - 89	(+)	0.938	ctcttTTCTctgaaatgct
V$STAT/STAT5.01	STAT5: signal transducer and activator of transcription 5	71 - 89	(-)	0.964	agcatTTCAgagaaaagag
V$GATA/LMO2COM.02	complex of Lmo2 bound to Tal-1, E2A proteins, and GATA-1, half-site 2	94 - 106	(-)	0.994	gacaGATAcgggc
V$SRFF/SRF.01	Serum response factor	92 - 110	(+)	0.748	ctgcccgTATCtgtcactc
V$MEIS/MEIS1.01	Binding site for monomeric Meis1 homeodomain protein	100 - 108	(-)	0.971	gtGACAgat
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	109 - 121	(+)	0.964	tcggGAGGggaga
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	113 - 119	(+)	0.985	gaGGGGa
V$SORY/HMGIY.01	HMGI(Y) high-mobility-group protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex	114 - 130	(-)	0.922	ctggagAATTctcccct
V$SORY/HMGIY.01	HMGI(Y) high-mobility-group protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex	115 - 131	(+)	0.924	ggggagAATTctccaga
V$AP4R/TAL1BETAITF2.01	Tal-1beta/ITF-2 heterodimer	123 - 139	(+)	0.854	ttctcCAGAtgtctcta
V$RP58/RP58.01	Zinc finger protein RP58 (ZNF238), associated preferentially with heterochromatin	125 - 137	(-)	0.941	gagaCATCtggag
V$HAND/HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12	127 - 141	(-)	0.776	cttagaGACAtctgg
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	165 - 171	(+)	0.985	gaGGGGa
V$P53F/P53.02	Tumor suppressor p53 (5' half site)	181 - 201	(-)	0.924	ttaggcctcttcaCATGccca
V$AREB/AREB6.02	AREB6 (Atp1a1 regulatory element binding factor 6)	205 - 217	(-)	0.978	ctctCACCtgagc
V$HEN1/HEN1.01	HEN1	201 - 221	(+)	0.837	aggggctCAGGtgagagatgg
V$HIFF/DEC1.01	Basic helix-loop-helix protein known as Dec1, Stra13 or Sharp2	204 - 218	(-)	0.860	tctctCACCtgagcc
V$EBOR/DELTAEF1.01	deltaEF1	205 - 219	(+)	1.000	gctcAGGTgagagat
V$RXRF/VDR_RXR.01	VDR/RXR Vitamin D receptor RXR heterodimer site	209 - 225	(+)	0.868	aggtgagaGATGgcgac
V$BNCF/BNC.01	Basonuclin, cooperates with USF1 in rDNA PolI transcription)	232 - 250	(-)	0.880	ggggtctggcTGTCctcac
V$GREF/GRE.01	Glucocorticoid receptor, C2C2 zinc finger protein binds glucocorticoid dependent to GREs	233 - 251	(-)	0.890	cggggtctggctGTCCtca
V$AP4R/TH1E47.01	Thing1/E47 heterodimer, TH1 bHLH member specific expression	235 - 251	(+)	0.965	aggacagCCAGaccccg

	in a variety of embryonic tissues				
V$SMAD/SMAD3.01	Smad3 transcription factor involved in TGF-beta signaling	239 - 247	(-)	0.994	GTCTggctg
V$AP2F/AP2.01	Activator protein 2	246 - 258	(+)	0.897	acCCCGcaggaga
V$AP4R/TAL1BETAE47.01	Tal-1beta/E47 heterodimer	256 - 272	(+)	0.953	agaagCAGAtggtcctc
V$NEUR/NEUROG.01	Neurogenin 1 and 3 (ngn1/3) binding sites	258 - 270	(-)	0.973	ggaCCATctgctt
V$HAND/HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12	260 - 274	(-)	0.795	cagaggACCAtctgc
V$EVI1/EVI1.02	Ecotropic viral integration site 1 encoded factor	268 - 284	(+)	0.903	tcctctgasAAGAcaaa
V$EVI1/EVI1.01	Ecotropic viral integration site 1 encoded factor	273 - 289	(+)	0.736	tgasAAGAcaaaggaag
V$ETSF/NRF2.01	Nuclear respiratory factor 2	279 - 295	(+)	0.877	gacaaaGGAAgagatgc
V$P53F/P53.01	Tumor suppressor p53	287 - 307	(-)	0.662	gacCTTGgccctgcatctctt
V$SF1F/FTF.01	Alpha (1)-fetoprotein transcription factor (FTF), liver receptor homologue-1 (LHR-1)	297 - 309	(+)	0.952	gggcCAAGgtctg
V$RORA/NBRE.01	Monomers of the nur subfamily of nuclear receptors (nur77, nurr1, nor-1)	297 - 313	(+)	0.895	gggccAAGGtctggaga
V$NRSF/NRSF.01	Neuron-restrictive silencer factor	308 - 328	(+)	0.738	tggAGAAccgaggtcaggggt
V$RORA/RORA1.01	RAR-related orphan receptor alpha1	312 - 328	(+)	0.936	gaaccgaGGTCaggggt
V$SRFF/SRF.01	Serum response factor	330 - 348	(-)	0.748	gtggccaTATCtgccaggc
V$GATA/LMO2COM.02	complex of Lmo2 bound to Tal-1, E2A proteins, and GATA-1, half-site 2	334 - 346	(+)	0.989	ggcaGATAtggcc
V$SRFF/SRF.03	Serum responsive factor	331 - 349	(+)	0.797	cctggcagatATGGccaca
V$AREB/AREB6.01	AREB6 (Atp1a1 regulatory element binding factor 6)	345 - 357	(-)	0.943	cctctACCTgtgg
V$MEIS/MEIS1.01	Binding site for monomeric Meis1 homeodomain protein	373 - 381	(+)	0.960	gtGACAgga
V$STAT/STAT.01	Signal transducers and activators of transcription	380 - 398	(+)	0.962	gaggcttcccGGAAgcagg
V$STAT/STAT.01	Signal transducers and activators of transcription	380 - 398	(-)	1.000	cctgcttccgGGAAgcctc
V$ETSF/CETS1P54.01	c-Ets-1(p54)	384 - 400	(+)	0.973	cttccCGGAagcaggga
V$IKRS/IK3.01	Ikaros 3, potential regulator of lymphocyte differentiation	393 - 405	(+)	0.857	agcagGGAAcaca
V$PAX5/PAX9.01	Zebrafish PAX9 binding sites	396 - 424	(+)	0.797	agggaacaCACTgaggggtgtttgggatt
V$RREB/RREB1.01	Ras-responsive element binding protein 1	408 - 422	(-)	0.879	tCCCAaacacccctc
V$IKRS/IK2.01	Ikaros 2, potential regulator of lymphocyte differentiation	415 - 427	(+)	0.983	gtttGGGAttctg
V$STAT/STAT6.01	STAT6: signal transducer and activator of transcription 6	419 - 437	(-)	0.848	gctccTCCCtcagaatccc
V$NRSF/NRSE.01	Neural-restrictive-silencer-element	429 - 449	(+)	0.687	gggaggagcaCGGGgacgccc
V$NFKB/NFKAPPAB50.01	NF-kappaB (p50)	438 - 452	(+)	0.891	acGGGGacgccctgg
V$WHZF/WHN.01	Winged helix protein, involved in hair keratinization and thymus epithelium differentiation	441 - 451	(+)	0.962	gggACGCcctg
V$P53F/P53.03	Tumor suppressor p53 (3' half site)	453 - 473	(-)	0.928	gccctggcatggcCATGtctc
V$MOKF/MOK2.01	Ribonucleoprotein associated zinc finger protein MOK-2 (mouse)	460 - 480	(+)	0.794	ggccatgccagggCCCTgagg
V$MOKF/MOK2.01	Ribonucleoprotein associated zinc finger protein MOK-2 (mouse)	465 - 485	(-)	0.761	ccactcctcagggCCCTggca
V$NOLF/OLF1.01	Olfactory neuron-specific factor	465 - 487	(-)	0.836	tcccacTCCTcagggccctggca

V$EGRF/WT1.01	Wilms Tumor Suppressor	478 - 492	(+)	0.915	aggagTGGGagagcc
V$BCL6/BCL6.01	POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma	560 - 576	(-)	0.903	agaTTCCtagagactgc
V$AP4R/TH1E47.01	Thing1/E47 heterodimer, TH1 bHLH member specific expression in a variety of embryonic tissues	569 - 585	(-)	0.932	gcttcttCCAGattcct
V$ETSF/ETS2.01	c-Ets-2 binding site	581 - 597	(+)	0.905	gaagcAGGAattctttg
V$LEFF/LEF1.02	TCF/LEF-1, involved in the Wnt signal transduction pathway	588 - 604	(-)	0.954	gtatcctCAAAgaattc
V$ZBPF/ZF9.01	Core promoter-binding protein (CPBP) with 3 Krueppel-type zinc fingers	619 - 641	(-)	0.871	gaagtccCCACacctcagcctcc
V$MINI/MUSCLE_INI.01	Muscle Initiator Sequence	622 - 640	(-)	0.873	aagtccCCACacctcagcc
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	631 - 637	(+)	1.000	gtGGGGa
V$MINI/MUSCLE_INI.02	Muscle Initiator Sequence	634 - 652	(+)	0.879	gggactTCATgcagaagtc
V$TEAF/TEF1.01	TEF-1 related muscle factor	660 - 672	(+)	0.916	caCATTcccttgg
V$SF1F/FTF.01	Alpha (1)-fetoprotein transcription factor (FTF), liver receptor homologue-1 (LHR-1)	663 - 675	(-)	0.941	cttcCAAGgaat
V$STAT/STAT6.01	STAT6: signal transducer and activator of transcription 6	660 - 678	(+)	0.854	cacatTCCCttggaagccg
V$SMAD/SMAD4.01	Smad4 transcription factor involved in TGF-beta signaling	674 - 682	(-)	0.969	GTCTcggct
V$IRFF/IRF4.01	Interferon regulatory factor (IRF)-related protein (NF-EM5, PIP, LSIRF, ICSAT)	676 - 694	(+)	0.960	ccgagactGAAAccagcag
V$MYOD/MYF5.01	Myf5 myogenic bHLH protein	685 - 699	(+)	0.920	aaacCAGCagcagag
V$ETSF/PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells	703 - 719	(-)	0.891	ctgacaGGAActcacca
V$MEIS/MEIS1.01	Binding site for monomeric Meis1 homeodomain protein	711 - 719	(-)	0.951	ctGACAgga
V$TALE/TGIF.01	TG-interacting factor belonging to TALE class of homeodomain factors	714 - 720	(+)	1.000	tGTCAga
V$BARB/BARBIE.01	Barbiturate-inducible element	720 - 734	(+)	0.916	agtgAAAGgagaagg
V$MOKF/MOK2.02	Ribonucleoprotein associated zinc finger protein MOK-2 (human)	727 - 747	(-)	0.984	ccaccatggcgggCCTTctcc
V$YY1F/YY1.01	Yin and Yang 1	734 - 752	(+)	0.848	gcccgCCATggtgggtttg
V$OCT1/OCT1.06	Octamer-binding factor 1	747 - 761	(+)	0.802	ggtttgtgAATTccc
V$FAST/FAST1.01	FAST-1 SMAD interacting protein	748 - 762	(+)	0.880	gtttgtgAATTccca
V$NFKB/NFKAPPAB.01	NF-kappaB	749 - 763	(-)	0.894	ctGGGAattcacaaa
V$SORY/HMGIY.01	HMGI(Y) high-mobility-group protein I (Y), architectural transcription factor organizing the framework of a nuclear protein-DNA transcriptional complex	748 - 764	(-)	0.920	gctgggAATTcacaaac
V$IKRS/IK1.01	Ikaros 1, potential regulator of lymphocyte differentiation	753 - 765	(-)	0.925	ggctGGGAattca
V$GABF/GAGA.01	GAGA-Box	757 - 781	(-)	0.797	gagggAGAGgaagccaggctgggaa
V$ETSF/PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells	763 - 779	(-)	0.874	gggagaGGAAgccaggc
V$GABF/GAGA.01	GAGA-Box	761 - 785	(-)	0.790	cccagAGGGagaggaagccaggctg
V$SRFF/SRF.03	Serum responsive factor	817 - 835	(-)	0.793	gaggccctacACGGgctgt
V$AP4R/AP4.01	Activator protein 4	830 - 846	(-)	0.970	aagggCAGCtggaggcc
V$MYOD/MYF5.01	Myf5 myogenic bHLH protein	831 - 845	(-)	0.917	agggCAGCtggaggc

V$ETSF/GABP.01	GABP: GA binding protein	853 - 869	(-)	0.865	ctggagGGAAgaggagc
V$NRSF/NRSF.01	Neuron-restrictive silencer factor	886 - 906	(+)	0.734	ctcAGCAcccagcgcggtgtc
V$E2TF/E2.01	BPV bovine papilloma virus regulator E2	889 - 905	(-)	0.894	acaccgcgctGGGTgct
V$E2TF/E2.01	BPV bovine papilloma virus regulator E2	890 - 906	(+)	0.935	gcacccagcgCGGTgtc
V$CDEF/CDE.01	Cell cycle-dependent element, CDF-1 binding site (CDE/CHR tandem elements regulate cell cycle dependent repression)	894 - 906	(+)	0.940	ccagCGCGgtgtc
V$NKXH/NKX31.01	Prostate-specific homeodomain protein NKX3.1	907 - 919	(+)	0.847	gtctAAGTtttct
V$MYT1/MYT1.02	MyT1 zinc finger transcription factor involved in primary neurogenesis	908 - 920	(+)	0.892	tctAAGTtttctc
V$E2FF/E2F.01	E2F, involved in cell cycle regulation, interacts with Rb p107 protein	911 - 925	(-)	0.763	atggagaGAAAactt
V$NKXH/NKX25.02	Homeo domain factor Nkx-2.5/Csx, tinman homolog low affinity sites	918 - 930	(-)	0.903	tctTAATggagag
V$HOXF/HOX1-3.01	Hox-1.3, vertebrate homeobox protein	918 - 934	(+)	0.841	ctctccATTAagaactc
V$PDX1/ISL1.01	Pancreatic and intestinal lim-homeodomain factor	916 - 936	(-)	0.826	ctgagttctTAATggagagaa
V$MYT1/MYT1L.01	Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1	930 - 942	(-)	0.949	agaaAGCTgagtt
V$SP1F/GC.01	GC box elements	944 - 958	(-)	0.910	gaatgGGAGgagctt
V$TEAF/TEF1.01	TEF-1 related muscle factor	952 - 964	(+)	0.869	ccCATTcctagtt
V$AREB/AREB6.01	AREB6 (Atp1a1 regulatory element binding factor 6)	968 - 980	(+)	0.949	cccctACCTgagc
V$HOXT/MEIS1_HOXA9.01	Homeobox protein MEIS1 binding site	984 - 996	(-)	0.824	cTGATttctgaac
V$GFI1/GFI1.01	Growth factor independence 1 zinc finger protein acts as transcriptional repressor	988 - 1002	(+)	0.973	agaAATCagagagta
V$CREB/TAXCREB.01	Tax/CREB complex	998 - 1018	(-)	0.825	gggggaTGACttctattactc
V$ETSF/PU1.01	Pu.1 (Pu120) Ets-like transcription factor identified in lymphoid B-cells	1021 - 1037	(+)	0.869	agaaaaGGAAtttgtcc
V$IRFF/IRF7.01	Interferon regulatory factor 7 (IRF-7)	1020 - 1038	(+)	0.907	aaGAAAaggaatttgtccc
V$HMTB/MTBF.01	Muscle-specific Mt binding site	1026 - 1034	(+)	0.931	aggaATTTg
V$RREB/RREB1.01	Ras-responsive element binding protein 1	1037 - 1051	(+)	0.837	cCCCAaagaaacaga
V$AREB/AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	1041 - 1053	(-)	0.997	gttctGTTTcttt
V$P53F/P53.02	Tumor suppressor p53 (5' half site)	1040 - 1060	(+)	0.924	caaagaaacagaaCTTGtccc
V$IRFF/IRF7.01	Interferon regulatory factor 7 (IRF-7)	1042 - 1060	(+)	0.888	aaGAAAcagaacttgtccc
V$P53F/P53.03	Tumor suppressor p53 (3' half site)	1049 - 1069	(-)	0.947	tttctttgggggaCAAGttct
V$CLOX/CDPCR3.01	Cut-like homeodomain protein	1057 - 1073	(+)	0.755	tcccccaaagaaATGGa
V$IRFF/IRF3.01	Interferon regulatory factor 3 (IRF-3)	1064 - 1082	(+)	0.950	aagaaatgGAAAcaatggg
V$OCT1/OCT1.04	Octamer-binding factor 1	1066 - 1080	(+)	0.813	gaAATGgaaacaatg
V$FKHD/FKHRL1.01	Fkh-domain factor FKHRL1 (FOXO)	1066 - 1082	(+)	0.876	gaaatggaAACAatggg
V$SORY/SOX5.01	Sox-5	1071 - 1087	(+)	0.986	ggaaaCAATgggaaatg
V$RBPF/RBPJK.02	Mammalian transcriptional	1075 - 1089	(+)	0.951	acaaTGGGaaatggg

	repressor RBP-Jkappa/CBF1				
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	1085 - 1097	(+)	0.910	atggGAGGcgggg
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	1086 - 1100	(+)	0.899	tgggaGGCGgggga
V$MAZF/MAZR.01	MYC-associated zinc finger protein related transcription factor	1090 - 1102	(+)	0.883	aggcggGGGGacc
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	1094 - 1100	(+)	0.990	ggGGGGa
V$RREB/RREB1.01	Ras-responsive element binding protein 1	1093 - 1107	(-)	0.810	cCCCAggtcccccg
V$RXRF/FXRE.01	Farnesoid X - activated receptor (RXR/FXR dimer)	1136 - 1152	(+)	0.808	cggtaacTGGCccagca
V$REBV/EBVR.01	Epstein-Barr virus transcription factor R	1155 - 1175	(-)	0.854	ccgcccggatcctgaGGTGgg
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	1166 - 1180	(+)	0.920	atccgGGCGgggagg
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	1168 - 1180	(+)	0.904	ccggGCGGggagg
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	1172 - 1178	(+)	0.985	gcGGGGa
V$EKLF/EKLF.01	Erythroid krueppel like factor (EKLF)	1172 - 1184	(+)	0.925	gcggggaGGGTag
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	1173 - 1185	(+)	0.911	cgggGAGGgtagg
V$ETSF/ETS1.01	c-Ets-1 binding site	1178 - 1194	(+)	0.934	agggtAGGAagtatccc
V$EGRF/EGR1.01	Egr-1/Krox-24/NGFI-A immediate-early gene product	1197 - 1211	(+)	0.840	atgcctgGGTGtccc
V$BNCF/BNC.01	Basonuclin, cooperates with USF1 in rDNA PolI transcription)	1196 - 1214	(+)	0.890	gatgcctgggTGTCcccaa
V$NFKB/NFKAPPAB50.01	NF-kappaB (p50)	1200 - 1214	(-)	0.888	ttGGGGacacccagg
V$MYT1/MYT1L.01	Myelin transcription factor 1-like, neuronal C2HC zinc finger factor 1	1208 - 1220	(-)	0.986	ggaaAGTTgggga
V$EBOX/ATF6.01	Member of b-zip family, induced by ER damage/stress, binds to the ERSE in association with NF-Y	1224 - 1238	(+)	0.937	ccgCCACccccgcta
V$EGRF/NGFIC.01	Nerve growth factor-induced protein C	1224 - 1238	(-)	0.863	taGCGGgggtggcgg
V$ZBPF/ZBP89.01	Zinc finger transcription factor ZBP-89	1220 - 1242	(+)	0.946	caaaccgccaCCCCcgctatgga
V$SRFF/SRF.03	Serum responsive factor	1228 - 1246	(+)	0.794	cacccccgctATGGagatg
V$IRFF/ISRE.01	Interferon-stimulated response element	1238 - 1256	(+)	0.838	atggagatGAAActgagac
V$IRFF/IRF3.01	Interferon regulatory factor 3 (IRF-3)	1244 - 1262	(+)	0.860	atgaaactGAGAcagaagg
V$AP2F/AP2.01	Activator protein 2	1268 - 1280	(+)	0.910	ggCCCGctgccgc
V$ETSF/ETS1.01	c-Ets-1 binding site	1274 - 1290	(-)	0.921	tctggAGGAagcggcag
V$RP58/RP58.01	Zinc finger protein RP58 (ZNF238), associated preferentially with heterochromatin	1284 - 1296	(-)	0.850	agctCATCtggag
V$ECAT/NFY.03	Nuclear factor Y (Y-box binding factor)	1292 - 1306	(-)	0.812	gaaaCCCAtgagctc
V$PCAT/ACAAT.01	Avian C-type LTR CCAAT box	1307 - 1317	(+)	0.834	tccaCCAAgga
V$ETSF/ETS1.01	c-Ets-1 binding site	1309 - 1325	(+)	0.942	caccaAGGAagttttcc
V$MYT1/MYT1.02	MyT1 zinc finger transcription factor involved in primary neurogenesis	1314 - 1326	(+)	0.895	aggAAGTtttccg
V$E2FF/E2F.02	E2F, involved in cell cycle	1316 - 1330	(-)	0.884	ccagcggAAAActtc

	regulation, interacts with Rb p107 protein				
V$MZF1/MZF1.01	Myeloid zinc finger protein MZF1	1346 - 1352	(-)	0.985	gcGGGGa
V$MAZF/MAZ.01	Myc associated zinc finger protein (MAZ)	1344 - 1356	(-)	0.913	gaggGCGGggaga
V$SP1F/SP1.01	Stimulating protein 1 SP1, ubiquitous zinc finger transcription factor	1344 - 1358	(-)	0.942	aagagGGCGgggaga
V$TBPF/TATA.01	cellular and viral TATA box elements	1369 - 1385	(+)	0.966	gcgtaTAAAtgcagctg
V$MYOD/E47.01	MyoD/E47 and MyoD/E12 dimers	1376 - 1390	(+)	0.922	aatGCAGctgtttgc
V$RBPF/RBPJK.01	Mammalian transcriptional repressor RBP-Jkappa/CBF1	1404 - 1418	(-)	0.862	actcTGGGagcttct
V$GLIF/GLI1.01	Zinc finger transcription factor GLI1	1430 - 1444	(+)	0.895	gggaccaGCCAggag
V$HAML/AML3.01	Runt-related transcription factor 2 / CBFA1 (core-binding factor, runt domain, alpha subunit 1)	1447 - 1461	(-)	0.846	tggaGTGGcttgtct
V$BCL6/BCL6.02	POZ/zinc finger protein, transcriptional repressor, translocations observed in diffuse large cell lymphoma	1463 - 1479	(+)	0.772	gacccccTAGAaataac
V$MEF2/HMEF2.01	Myocyte enhancer factor	1462 - 1484	(+)	0.891	ggacccccctagAAATaacctctc
V$EKLF/BKLF.01	Basic krueppel-like factor (KLF3)	1487 - 1499	(-)	0.960	ggGGGTgtgtctt
V$GABF/GAGA.01	GAGA-Box	1510 - 1534	(-)	0.819	gtgtgAGAGggagagagtcgtccgg
V$GABF/GAGA.01	GAGA-Box	1512 - 1536	(-)	0.794	gcgtgTGAGagggagagagtcgtcc
V$AHRR/AHR.01	Aryl hydrocarbon / dioxin receptor	1524 - 1546	(-)	0.805	ccctggggcaGCGTgtgagaggg
V$AP4R/AP4.02	Activator protein 4	1555 - 1571	(+)	0.970	tctccCAGCtggacctg
V$HAND/HAND2_E12.01	Heterodimers of the bHLH transcription factors HAND2 (Thing2) and E12	1559 - 1573	(-)	0.778	ctcaggTCCAgctgg
V$PAX6/PAX6.02	PAX6 paired domain and homeodomain are required for binding to this site	1557 - 1575	(-)	0.908	ggctcaggtCCAGctggga

Appendix 3

TNFα promoter data set for the suid phylogeny study. Includes two bearded pigs (*Sus barbatus subspp.*), an additional babirusa, two further warthogs, two African bushpigs (*Potamochoerus africanus*), two red river hogs (*Potamochoerus porcus*), a Javan warty pig (*Sus verrrucosus*), three Sulawesi warty pigs (*Sus celebensis*) and a white-lipped peccary (*Tayassu peccari*). Two extra domestic pig sequences were also retrieved from the public sequence database (accession numbers in brackets). The insertion seen in the Javan warty pig sequence has been highlighted.

>Sus scrofa (1724bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTGA
GAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCGAGGTCGGGGGTCGCCTGGCAGATATGG
CCACAGGTAGAGGGACAGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGGG
TGTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCATGAGGAG
TGGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCAGGAG
CGCCGAGGGCCTGGACAGTCTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACAC
AAAGGGAGGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACT
GAAACCAGCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTG
TGAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGC
CCGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAG
CACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCC
TAGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGG
AATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGG
GGGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGG
ATCCGGGCAGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCC
GCTATGGAGATGAAACTAAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGT
TTCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGT
ATAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAG
GAGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGC
CGGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTC
TGAAAAAGACACC
>Sus scrofa (AJ251914) (1724bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTGA
GAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCGAGGTCGGGGGTCGCCTGGCAGATATGA
CCACAGGTAGAGGGACAGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGGG
TGTTTGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCATGAGGAG
TGGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCAGGAG
CGCCGAGGGCCTGGACAGTCTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACAC
AAAGGGAGGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACT
GAAACCAGCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTG
TGAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGC
CCGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAG
CACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCC
TAGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGG
AATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGG
GGGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGG
ATCCGGGCAGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCC
GCTATGGAGATGAAACTAAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGT
TTCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGT
ATAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAG
GAGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGC
CGGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTC
TGAAAAAGACACC
>Sus scrofa (X54859) (1724bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTTCCTCTG
AGAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCGAGGTCGGGGGTCGCCTGGCAGATATG
GCCACAGGTAGAGGGACAGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGG
GTGTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGCCGTCCAGGGCCATGAGGAGT
GGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCAGGAGC
CGCGAGGGCCTGGACAGTCTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACA
AAGGGAGGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACTG

AAACCAGCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGT
GAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGCC
CGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGC
ACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCT
AGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGGA
ATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGGG
GGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGGA
TCCGGGCAGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCCG
CTATGGAGATGAAACTAAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTT
TCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTA
TAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGG
AGAGAGACAAGCCATCTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGC
CGGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTC
TGAAAAAGACACC

>Sus barbatus oi (1724bp)

GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTGA
GAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCGAGGTCGGGGGTCGCCTGGCAGATATGG
CCACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGGG
TGTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCATGAGGAG
TGGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCCGGAG
CGCCGAGGGCCTGGACAGTTTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACTGTGGCCACA
CAAAGGGAGGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGAC
TGAAACCAGCAGCAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGT
GAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGCC
CGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGC
ACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCT
AGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGGA
ATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGGG
GGGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGG
ATCCGGGCAGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCC
GCTATGGAGATGAAACTAAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGT
TTCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGT
ATAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAG
GAGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGC
CGGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTC
TGAAAAAGACACC

>Sus barbatus barbatus (1723bp)

GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTGA
GAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCAAGGTCGGGGGTCGCCTGGCAGATATGG
CCACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGGG
TGTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCATGAGGAG
TGGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCAGGAG
CGCCGAGGGCCTGGACAGTCTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACAC
AAAGGGAGGCTGAGATGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACT
GAAACCAGCAGCAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTG
AATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGCCC
GTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCA
CCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCTA
GTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGGAA
TTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGGGG
GGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGGA
TCCGGGCAGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCCG
CTATGGAGATGAAACTAAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTT
TCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTA
TAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGG

AGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAACCTGTCAGAAGACACACCCCCGAACAGGCAGCC
GGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTCT
GAAAAAGACACC
>Babyrousa babyrussa 1(1723bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACAGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGCGTGCAAAGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCGGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTCA
GAAGACAAAGGAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATACGGC
CACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGCAGGGAACACACTGAGGGGT
GTTCGGGATTCTGAGGGAGGAGCACGGGGATGCCCTGGGATATATGGCCGTGCCAGGGCCCTGAGGAGT
GGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCACGCCCAGGAGC
ACTGAGGGCCTGGACAGTCTCCAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATATGTGGCCACACA
AAGGGAAGCTGAGTTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCAGAGACTG
AAACCAGCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGT
GAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGCC
CGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGC
ACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGGACTCAGCTTTCTGAAGCTCCTCCCATTCCT
AGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGGA
ATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGGG
GCAACTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGGA
TCCGGGCGGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCCG
CTATGGAGATGAAACTGAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTT
TCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTA
TAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGG
AGAGAGACAAGCCACTCCAGAACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGCC
GGACGACTCTTTACCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTCT
GAAAAAGACACC
>Babyrousa babyrussa 2 (1723bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACAGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGCGTGCAAAGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCGGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTCA
GAAGACAAAGGAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATACCGC
CACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGCAGGGAACACACTGAGGGGT
GTTCGGGATTCTGAGGGAGGAGCACCGGGATGCCCTGGGAGATATGGCCGTGCCAGGGCCCTGAGGAGT
GGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCACGCCCAGGAGC
ACTGAGGGCCTGGACAGTCTCCAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACA
AAGGGAAGCTGAGTTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCAGAGACTG
AAACCAGCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGT
GAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGCC
CGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGC
ACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGGACTCAGCTTTCTGAAGCTCCTCCCATTCCT
AGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGGA
ATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGGG
GCAACTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGGA
TCCGGGCGGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCCG
CTATGGAGATGAAACTGAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTT
TCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTA
TAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGG
AGAGAGACAAGCCACTCCAGAACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGCC
GGACGACTCTCTACCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTCT
GAAAAAGACACC
>Phacochoerus africanus 1 (1716bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACAGGAGCTGTGGGCATGTGAAGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGACCTCGGGGTGAGGACAGCCAGACCCCGCAGGAGAAGCAGATGGTCCTCTGAS
AAGACAAAGGAAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATATGGC
CACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGAAGCAGGGAACACACTGAGGGGT
GTTTGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCATGCCAGGGCCCTGAGGAGT

GGGAGAGCCTCTGAGGCTAGCTGCTGGAGATACAGGGACATTTGAGGAGACACGGTCAGGAGCGCCGAG
GGCCTGGGCAGTCTCTAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACAAAGGGA
GGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCCCACATTCCCTTGGAAGCCGAGACTGAAACCA
GCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTGAATTC
CCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCACAGCCCGTGTA
GGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCACCCAG
CGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCTAGTTCT
ACCCCTACCTGAGCCCTGTTCAGAAATCAGAGAGTAATAGAAGTCATCCCCCAAAGAAAAGGAATTTGT
CCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCGGGGGGACCTG
GGGTCCAGCCTCCAGGGTCCTACACGCAGAGCGGTAACTGGCCCAGCAAGCCCACCTCAGGATCCGGGC
GGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCACCCCCGCTATGGA
GATGAAACTGAGACAGAAGGTGCAGGGCCCGCTGCCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCAC
CAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGC
AGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGGAGAGAGA
CAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGCCGGACGAC
TCTCTCCCTCTCACACGCTGCCCCAGGGCACCACCATCTCCCAGCTGGACCTGAGCCCCTCTGAAAAAG
ACACC

> Phacochoerus africanus 2 (1714bp)

GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACSGCTGACCCTCAATGAAACCCAATAAA
GCTCTTTTATCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACAGGAGCTGTGGGCATGTGAAGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGACCTCGGGGTGAGGACAGCCAGACCCCGCAGGAGAAGCAGATGGTCCTCTGAG
AAGACAAAGGAAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATATGGC
CACAGGTAGAGGGACGGAGGAATAGGTGACAGGAGGCTTCCCGGAAGCAGGGAACACACTGAGGGGTGT
TCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCCTGAGGAGTGG
GAGAGCCTCTGAGGCTAGCTGCTGGAGATACAGGGACATTTGAGGAGACACGGTCAGGAGCGCCGAGGG
CCTGGGCAGTCTCTAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACAAAGGGAGG
CTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACTGAAACCAGC
AGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTGAATTCCC
AGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCACAGCCCGTGTAGG
GCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCACCCAGCG
CGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCCAGTTCTAC
CCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGGAATTTGTCC
CCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCGGGGGGACCTGGG
GTCCAGCCTCCAGGGTCCTACACGCAGAGCGGTAACTGGCCCAGCAAGCCCACCTCAGGATCCGGGCGG
GGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCACCCCCGCTATGGAGA
TGAAACTGAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCACCA
AGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGCAG
CTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGGAGAGAGACA
AGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGCCGGACGACTC
TCTCCCTCTCACACGCTGCCCCAGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTCTGAAAAAGAC
ACC

> Phacochoerus africanus 3 (1483bp)

TCTCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCTAAGGAGT
GGAGGGAGGACAGGAATCAGAGGGGACAGGAGCTGTGGGCATGTGAAGAGGCCTAAGGGGCTCAGGTGA
GAGATGGCGACCTCGGGGTGAGGACAGCCAGACCCCGCAGGAGAAGCAGATGGTCCTCTGAGAAGACAA
AGGAAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATATGGCCACAGGT
AGAGGGACGGAGGAATAGGTGACAGGAGGCTTCCCGGAAGCAGGGAACACACTGAGGGGTGTTCGGGAT
TCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCCTGAGGAGTGGGAGAGCC
TCTGAGGCTAGCTGCTGGAGATACAGGGACATTTGAGGAGACACGGTCAGGAGCGCCGAGGGCCTGGGC
AGTCTCTAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACAAGGGGAGGCTGAGGT
GTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACTGAAACCAGCAGCAGAG
TTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTGAATTCCCAGCCTGG
CTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCACAGCCCGTGTAGGGCCTCCA
GCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCACCCAGCGCGGTGTC
GTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCTAGTTCTACCCCTACC
TGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGGAATTTGTCCCCCAAAG
AAATGGAAACAATGGGAAATGGGAGGCGGGGGGACCTGGGGTCCAGCCTCCAGGGTCCTACACGCAGAG
CGGTAACTGGCCCAGCAAGCCCACCTCAGGATCCGGGCGGGGAGGGTAGGAAGTATCCCTGATGCCTGG
GTGTCCCCAACTTTCCAAACCGCCACCCCCGCTATGGAGATGAAACTGAGACAGAAGGTGCAGGGCCCG
CTACCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGC

CTCTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTC
CCAGAGTGAGGACACCAGGGGACCAGCCAGGAGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAACC
TCTCAGAAGACACACCCCCGAACAGGCAGCCGGACGACTCTCTCCCTCTCACACGCTGCCCCAGGGTGC
CACCATCTCCCAGCTGGACCTGAGCCCCTCTGAA
>Potamochoerus larvatus 1 (1716bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGATTTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGCATGTGAAGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCGGGGTGAGGGCAGCCAGACCCCGCAGGAGAAGCAGATGGTCCTCTGAG
AAGACAAAGGAAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATATGGC
CCCAGGTAGAGGGACGAAGGAATAGGGGTGACAGGAGGCTTCCCGGAAGCAGGGAACACACTGAGGGGT
GTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCCTGAGGAGT
GGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCAGGAGCGCCGAG
GGCCTGGGCAGTCTCTAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACAAAGGGA
GGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACTGAAACCA
GCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTGAATTC
CCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCACAGCCCGTGTA
GGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCACCCAG
CGCGGTGTCGTCTGAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCTAGTTCT
ACCCCTACCTGAGCCCTGTTTGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGAAATTTGT
CCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCGGGGGGACCTG
GGGTCCAGCCTCCAGGGTCCTACACGCAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGGATCCGGGC
GGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCCGCTATGGA
GATGAAACTGAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCAC
CAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGC
AGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGGAGAGAGA
CAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGCCGGATGAC
TCTCTCCCTCTCACACGCTGCCCCAGGTCGCCACCATCTCCCAGCTGGACCTGAGCCCCTCTGAAAAAG
ACACC
> Potamochoerus larvatus 2 (1716bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGCATGTGAAGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCGGGGTGAGGGCAGCCAGACCCCGCAGGAGAAGCAGATGGTCCTCTGAG
AAGACAAAGGAAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATATGGC
CCCAGGTAnAGGGACGAAGGAATAGGGGTGACAGGAGGCTTCCCGGAAGCAGGGAACACACTGAGGGGT
GTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGCGAGACATGGCCGTGCCAGGGCCCTGAGGAGT
GGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCAGGAGCGCCGAG
GGCCTGGGCAGTCTCTAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACAAAGGGA
GGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACTGAAACCA
GCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTGAATTC
CCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCACAGCCCGTGTA
GGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCACCCAG
CGCGGTGTCGTCTGAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCTAGTTCT
ACCCCTACCTGAGCCCTGTTTGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGAAATTTGT
CCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCGGGGGGACCTG
GGGTCCAGCCTCCAGGGTCCTACACGCAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGGATCCGGGC
GGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCCGCTATGGA
GATGAAACTGAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCAC
CAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGC
AGCTGTTTGCACACCCAGCCAGCAGAGGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGGAGAGAGA
CAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGCCGGACGAC
TCTCTCCCTCTCACACGCTGCCCCAGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTCTGAAAAAG
ACACC
>Potamochoerus porcus 1 (1716bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGCATGTGAAGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCGGGGTGAGGGCAGCCAGACCCCGCAGGAGAAGCAGATGGTCCTCTGAG
AAGACAAAGGAAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATATGGC
CACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGAAGCAAGGAACACACTGAGGGGT

GTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCCTGAGGAGT
GGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCAGGAGCGCCGAG
GGCCTGGGCAGTCTCTAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACAAAGGGA
GGCTGAGGTGTGGGGACTTCATGCAGAAATCAGGGCCTCACATTCCCTTGGAAGCCGAGACTGAAACCA
GCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTGAATTC
CCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCACAGCCCGTGTA
GGGCCTTCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCACCCAG
CGCGGTGTCGTCTGAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCTAGTTCT
ACCCCTACCTGAGCCCTGTTTGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGAAATTTGT
CCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCGGGGGGACCTG
GGGTCCAGCCTCCAGGGTCCTACACGCAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGGATCCGGGC
GGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCCGCTATGGA
GATGAAACTGAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCAC
CAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGC
AGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGGAGAGAGA
CAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGCCGGACGAC
TCTCTCCCTCTCACACGCTGCCCCAGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTCTGAAAAAG
ACACC

> Potamochoerus porcus 2 (1716bp)

GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCCCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGCATGTGAAGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCGGGGTGAGGGCAGCCAGACCCCGCAGGAGAAGCAGATGGTCCTCTGAG
AAGACAAAGGAAGAGATGCAGGGCCAAGGTCTGGAGAACCGAGGTCAGGGGTCGCCTGGCAGATATGGC
CACAGGTATAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGAAGCAAGGAACACACCTGACGGG
TGTTCGGGATCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCCTGAGGAGT
GGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCAGGAGCGCCGAG
GGCCTGGGCAGTCTCTAGGAATCTGGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACACAAAGGGA
GGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACTGAAACCA
GCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTGAATTC
CCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCACAGCCCGTGTA
GGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCACCCAG
CGCGGTGTCGTCTGAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCCTAGTTCT
ACCCCTACCTGAGCCCTGTTTGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGAAATTTGT
CCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCGGGGGGACCTG
GGGTCCAGCCTCCAAGGTCCTACACGCAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGGATCCGGGC
GGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCCGCTATGGA
GATGAAACTGAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCAC
CAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGC
AGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGGAGAGAGA
CAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGCCGGACGAC
TCTCTCCCTCTCACACGCTGCCCCAGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTCTGAAAAAG
ACACC

>Sus verrucosus (2031bp)

GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTGA
GAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCGAGGTCGGGGGTCGCCTGGCAGATATGG
CCACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGGG
TGTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTGGGAGACATGGCCGTGCCAGGGCCATGAGGAG
TGGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCAGGAG
CGCCGAGGGCCTGGACAGTCTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACAC
AAAGGGAGGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACT
GAAACCAGCAGCAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTGTG
AATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGCCC
GTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAGCA
CCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGGGAGTTCCCGTCGTGGCG
CAGTGGTTAACGAATCCGACTAGGAACCATGAGGTTGCGGGTTCGGTCCCTGCCCTTGCTCAGTGGGTT
AACGATCCGGCGTTGCCCTGAGCTGTGGTGTAGGTTGCAGACGCGGCTCGGATTCTGAGTTGCTGTGGC
TCTGGCGTAGGCCGGTGGCTACAGCTCCGATTCGACCCCTAGCCTGGGAACCTCCATATGCCGCGGGAG

TGGCCCAAGAAATAGCAACAATAACAACAACAGCAAAAAAGACAAAAGACAAAAAAAAAAAAAAAAAA
GAACTCAGCTTTCTGAAGCTCCTCCCATTCCTAGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGA
GAGAAATAGAAGTCATCCCCCAAAGAAAAGGAATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAG
AAATGGAAACAATGGGAAATGGGAGGCAGGGGGGGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAG
AGCAGTAACTGGCCCAGCAAGCCCACCTCAGGATCCGGGCAGGGAGGGTAGGAAGTATCCCTGATGCCT
GGGTGTCCCCAACTTTCCAAACCGCCGCCCCCGCTATGGAGATGAAACTAAGACAGAAGGTGCAGGGCC
CGCTACCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGA
GCCTCTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGC
TCCCAGAGTGAGACACCAGGGGACCAGCCAGGAGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAAC
CTCTCAGAAGACACACCCCCGAACAGGCAGCCGGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCG
CCACCATCTCCCAGCTGGACCTGAGCCCCTCTGAAAAAGACACC
>Sus celebensis 1 (1724bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTGA
GAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCGAGGTCGGGGGTCGCCTGGCAGATATGG
CCACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGGG
TGTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTAAGAGACATGGCCGTGCCAGGGCCATGAGGAG
TGGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCAGGAG
CGCCGAGGGCCTGGACAGTCTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACAC
AAAGGGAGGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACT
GAAACCAGCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTG
TGAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGC
CCGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAG
CACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCC
TAGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGG
AATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGG
GGGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGG
ATCCGGGCGGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCC
GCCATGGAGATGAAACTAAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGT
TTCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGT
ATAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAG
GAGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGC
CGGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTC
TGAAAAAGACACC
>Sus celebensis 2 (1724bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGAAGGACAGGAATCAGAGGGGACGGGAGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTGA
GAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCGAGGTCGGGGGTCGCCTGGCAGATATGG
CCACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGGG
TGTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTAGGAGACATGGCCGTGCCAGGGCCATGAGGAG
TGGGAGAGCCTCTGAGGCTAGCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCAGGAG
CGCCGAGGGCCTGGACAGTCTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACAC
AAAGGGAGGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACT
GAAACCAGCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTG
TGAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGC
CCGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAG
CACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCC
TAGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGG
AATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGG
GGGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGG
ATCCGGGCGGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCC
GCCATGGAGATGAAACTAAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGT
TTCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGT
ATAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAG
GAGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGC
CGGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTC
TGAAAAAGACACC

>Sus celebensis 3 (1724bp)
GGCTCAGGAAGGGGCTGCTTGACTGGAGGCTCATGAGGAGACGGCTGACCCTCGATGAAACCCAATAAA
GCTCTTTTCTCTGAAATGCTGTCTGCTCGTATCTGTCACTCGGGAGGGGAGAATTCTCCAGATGTCTCT
AAGGAGTGGAGGGAGGACAGGAATCAGAGGGGACGGGGGCTGTGGGTGTGTGATGAGGCCTAAGGGGCT
CAGGTGAGAGATGGCGGCCTCAGGGTGAGGGCAGCCAGACCCCTGCAGGAGAAGCAGATGGTCCTCTGA
GAAGACAAAGGAAGAGATGCAGGGCCAAGGTCTTGAGAACCGAGGTCGGGGGTCGCCTGGCAGATATGG
CCACAGGTAGAGGGACGGAGGAATAGGGGTGACAGGAGGCTTCCCGGGAGAAGGGAACACACTGAGGGG
TGTTCGGGATTCTGAGGGAGGAGCACGGGGACGCCCTAAGAGACATGGCCGTGCCAGGGCCATGAGGAG
TGGGAGAGCCTCTGAGGCTATCGGCTGGAGATACAGGGACATTTGAGGAGACACGGTCATGGCCAGGAG
CGCCGAGGGCCTGGACAGTCTCTAGGAATCTCGAAGAAGCAGGAATTCTTTGAGGATACGTGGCCACAC
AAAGGGAGGCTGAGGTGTGGGGACTTCATGCAGAAGTCAGGGCCTCACATTCCCTTGGAAGCCGAGACT
GAAACCAGCAGCAGAGTTTTGGTGAGTTCCTGTCAGAGTGAAAGGAGAAGGCCCGCCATGGTGGGTTTG
TGAATTCCCAGCCTGGCTTCCTCTCCCTCTGGGGCTGTCCCAGGCCTGTTCCTGCCGTCCTCCCCCAGC
CCGTGTAGGGCCTCCAGCTGCCCTTCTCCCAGCTCCTCTTCCCTCCAGGAGACGAAACATGGGTCTCAG
CACCCAGCGCGGTGTCGTCTAAGTTTTCTCTCCATTAAGAACTCAGCTTTCTGAAGCTCCTCCCATTCC
TAGTTCTACCCCTACCTGAGCCCTGTTCGGAAATCAGAGAGAAATAGAAGTCATCCCCCAAAGAAAAGG
AATTTGTCCCCCAAAGAAACAGAACTTGTCCCCCAAAGAAATGGAAACAATGGGAAATGGGAGGCAGGG
GGGACCTGGGGTCCAGCCTCCAGGGTCCTACACACAGAGCAGTAACTGGCCCAGCAAGCCCACCTCAGG
ATCCGGGCGGGGAGGGTAGGAAGTATCCCTGATGCCTGGGTGTCCCCAACTTTCCAAACCGCCGCCCCC
GCCATGGAGATGAAACTAAGACAGAAGGTGCAGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGT
TTCTCCACCAAGGAAGTTTTCCGCTGGTTGAAAGAGAGCCTCTCCCCGCCCTCTTCTCACCCAGAGCGT
ATAAATGCAGCTGTTTGCACACCCAGCCAGCAGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAG
GAGAGAGACAAGCCACTCCAGGACCCCCTAGAAATAACCTCTCAGAAGACACACCCCCGAACAGGCAGC
CGGACGACTCTCTCCCTCTCACACGCTGCCCCGGGGCGCCACCATCTCCCAGCTGGACCTGAGCCCCTC
TGAAAAAGACACC
>Tayassu peccari (1692bp)
GGCTCAGAGCCAGACGCTGCTTGGCTGGACGCTCACGAGGAGATGACTGACCCTCGACGAAACCTAATA
AACCTGTTTTCTCTGAAATGCTGTCTGCCTGGATCTGTCGTTCGGGAGGGGGGACTTCCTTAGATGTCT
CTAAGGAATGGGGGGAGGGCAGGAATCAGGGCATGCGGGAGCTGTGGGCGAACAAAGAGGCCTAAGGGG
CTGATGTGAAATATGGCGGCCTCGGGGTGAGGGCGGCCTCGGGGTGAGGGCAGCCAGAGCCCCACAGGA
GAAGCAGAGGGTCTGCTGAGAAGACAAAGGAAAGATGCAGGGCCAAGCTCCGGAGCACTGGAGGTCAAG
GCCCCGGCAGACATGGCCACGAGTAGAGGGCCGGAGGAACGCGGGTGACGGAGGCTTCCGGGGAGCAGT
GAACACACTGAAGGGCATTCGGGATTCTGGGAGAGGAGCATGGGGACCCCCTGGGAGACACGGCTCCGC
CAGGGCCCTGAGGGGTGGAAGAGCCTCTGAGGCTAGTGGCTGGAGACACAGGGGCCTTTGAGGAGACAC
GGTCACGCCCAGGGGCTCTGAGGGCCACACAGTCTACAGGAATTTGGTAGAAGCAGGAATTCTTGGAGG
GTCTGTGGCCACATAAAGGGACTCTGAGGCAAGGGGACTTCGTGCAAAAGTCAGGGCCTCACATTCCCT
TGGAAGCAAAGACTGAAACCAGCAGCAGCGTTTCGGTGAGTTCCTGTCAGAGTGAAAGCAGAAGGCCCG
CCATGGTGGGGTCCGTGAATTCCCAGCCGGGCTTCCTCTCCCTCTGGGGCTGTCCCAGTCCTGTTCCTG
CAATTCCTCCCCCAGCCCGTGTTAGGGCCCCAAGCTCCCCTTCTCTCAGCTCCTCTTCCCCCCAGGGGA
CCAAACACAGGTCTCAGGACCCAGCTCAGCTTCCCCTAACTTTTCTCACCATTAAGGATTCAGTTTTCT
GAATCTCCTTCCATTTCCTGAGTCCCGTCTGGAAATCAGAAGGAACTAGAACTTGTCCCCCAAAGAAAT
GGAACCTGTCCCCCAAAGAAATGGAACTCGTCCCCCAAAGAAACAGAACCTGTCCCCCAAAGGGTCCTA
CACACAGAAGGGTCACTGGCCCAGCAAGCCCCCCTCGAATCCGGGTGGGGAGGATAGGAAGTATCCTTG
ATGCCTGGGTGTCCCCAACTTTCCAAACCGCCACCCCCGCGATGGGGATGAAACTGAGACAGAAGGTGC
AGGGCCCGCTACCGCTTCCTCCAGATGAGCTCATGGGTTTCTCCACCAAGGAAGTTTTCCGCTGGTTGA
ATGAGAGCCTTTCCCCGCCCTCTTCTCACCCAGAGCGTATAAATGCAGCTGTTTGCACACCCAGCCAGC
AGAAGCTCCCAGAGTGAGGACACCAGGGGACCAGCCAGGAGAGAGAGACAAGCCACTCCAGAACCCCCT
GGAAATAAGCTCTCAGAAGACACACCCCCGACAGGCAGCCAGACGACTCTCTCCCTCTAAGCACACGCT
GCCCGGGGGTGCCACCATCTCCCAGCTGGACCTGAGCCCCTCTGAAAAAGACACC